



**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO RIO
GRANDE DO NORTE**

DAMARIS DOS SANTOS TAVARES

**ANÁLISE DESCRITIVA E EXPLORATÓRIA SOBRE CURSOS FIC OFERTADOS
PELO IFRN**

Canguaretama, RN – 2026

DAMARIS DOS SANTOS TAVARES

**ANÁLISE DESCRITIVA E EXPLORATÓRIA SOBRE CURSOS FIC OFERTADOS
PELO IFRN**

(Ficha Catalográfica no verso, Biblioteca responsável pela elaboração)

Trabalho de Conclusão de Curso
apresentado ao Curso Superior de
Tecnologia em Sistemas para Internet do
Instituto Federal de Educação, Ciência e
Tecnologia do Rio Grande do Norte, em
cumprimento às exigências legais como
requisito parcial à obtenção do título de
Tecnólogo em Sistemas para Internet.

Orientador: Dr. Bruno Gomes de Araujo

CANGUARETAMA/RN

2026

DAMARIS DOS SANTOS TAVARES

**ANÁLISE DESCRITIVA E EXPLORATÓRIA SOBRE CURSOS FIC OFERTADOS
PELO IFRN**

Trabalho de Conclusão de Curso
apresentado ao Curso Superior de
Tecnologia em Sistemas para Internet do
Instituto Federal de Educação, Ciência e
Tecnologia do Rio Grande do Norte, em
cumprimento às exigências legais como
requisito parcial à obtenção do título de
Tecnólogo em Sistemas para Internet.

Aprovado em: 11/02/2026

Banca Examinadora



Documento assinado digitalmente

BRUNO GOMES DE ARAUJO

Data: 09/06/2026 13:40:31-0300

Verifique em <https://validar.iti.gov.br>

Dr. Bruno Gomes de Araujo

Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte



Documento assinado digitalmente

TATIANA AMARAL SORRENTINO

Data: 09/06/2026 15:29:05-0300

Verifique em <https://validar.iti.gov.br>

Dra. Tatiana Amaral Sorrentino - Examinadora

Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte



Documento assinado digitalmente

EDSON ANÍBAL DE MACEDO REIS BATISTA

Data: 16/06/2026 12:12:44-0300

Verifique em <https://validar.iti.gov.br>

Me. Edson Aníbal de Macedo Reis Batista - Examinador

Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte

RESUMO

Este trabalho realizou uma análise descritiva e exploratória sobre os cursos de Formação Inicial e Continuada (FIC) ofertados pelo IFRN, a partir de dados institucionais extraídos do SUAP no período de 2020 a 2026. Os dados foram exportados em planilhas anuais e consolidados em um único conjunto contendo 79.231 registros, permitindo a execução de etapas de preparação (padronização, tratamento de nulos e conversões de tipos) e, posteriormente, a geração de sínteses e visualizações para identificar padrões de oferta. Os resultados evidenciaram a predominância de cursos FIC- (carga horária inferior a 160 horas) ao longo da série, com intensificação recente dos cursos FIC+ (≥ 160 horas), especialmente a partir de 2024. Além disso, foram analisadas distribuições por campus, turno e um recorte específico do Campus Canguaretama, apoiando a interpretação dos dados e a organização de informações relevantes para subsidiar um dashboard voltada ao acompanhamento das ofertas.

Palavras-chaves: Sistema Unificado de Administração Pública(SUAP); Cursos; Formação Inicial e Continuada(FIC); Análise de Dados; Python.

LISTA DE ILUSTRAÇÕES

FIGURA 1. ETAPAS DE PREPARAÇÃO DE DADOS.	18
FIGURA 2. RESULTADO DA ANÁLISE NO GRÁFICO.	23
FIGURA 3. CURSOS FIC COM CARGA HORÁRIO INFERIOR A 160 HORAS POR ANO.	27
FIGURA 4. CURSOS FIC COM CARGA HORÁRIA IGUAL OU SUPERIOR A 160 HORAS POR ANO.	28
FIGURA 5. CINCO CAMPI QUE MAIS OFERTARAM CURSO DE 2020 A 2026.	29
FIGURA 6. CINCO CAMPI COM MENOR NÚMERO DE CURSOS OFERTADOS .	30
FIGURA 7. DISTRIBUIÇÃO ANUAL DAS OFERTAS POR TURNO.	31
FIGURA 8. OFERTA DE CURSOS FIC COM CARGA HORÁRIA INFERIOR A 160 HORAS POR ANO E TURNO .	32
FIGURA 9. CURSOS OFERTADOS COM CARGA HORÁRIA IGUAL OU SUPERIOR A 160 HORAS POR TURNO E ANO.	33
FIGURA 10. DISTRIBUIÇÃO DOS 22 CURSOS DO CAMPUS CANGUARETAMA ENTRE AS CATEGORIAS FIC- E FIC+.	35
FIGURA 11. CURSOS COM CARGA HORÁRIA INFERIOR A 160 HORAS NO CAMPUS CANGUARETAMA POR ANO	36
FIGURA 12. CURSOS COM CARGA HORÁRIA IGUAL OU SUPERIOR A 160 HORAS NO CAMPUS CANGUARETAMA POR ANO.	37

LISTA DE TABELAS

TABELA 1. QUANTIDADE DE LINHAS NOS DATASETS POR ANO.	17
TABELA 2. QUANTIDADE DE DADOS NULOS	19
TABELA 3. QUANTIDADE DE CURSOS OFERTADOS POR ANO CONFORME A CATEGORIA (FIC- E FIC+)	25

SUMÁRIO

1. INTRODUÇÃO	8
1.1 Objetivo Geral	10
1.2 Objetivos específicos	10
2. REFERENCIAL TEÓRICO	11
2.1. Análise de Dados	11
3. METODOLOGIA	15
4. RESULTADOS	16
4.1. <i>Dataset</i> analisado	16
4.2. Preparação dos Dados	17
4.2.1. Concatenação dos Datasets	18
4.2.2. Substituição de Caracteres textuais	18
4.2.3. Dados Nulos	19
4.2.4. Conversão de Tipos de Dados	21
4.2.5. Dados atípicos (<i>outliers</i>)	22
4.3. Análises Descritivas e Exploratórias	23
4.3.1. Modalidades	23
4.3.2. Informações Gerais	24
4.3.3. Ofertas	28
4.3.4. Turnos	30
4.3.5. Ofertas de Cursos FIC no Campus Canguaretama	33
5.CONCLUSÃO	37
REFERÊNCIAS	38
ANEXOS	39

1. INTRODUÇÃO

A oferta de cursos de Formação Inicial e Continuada (FIC) integra as ações estratégicas do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte, ao promover uma educação profissional e tecnológica comprometida com a formação humana integral e com as demandas sociais e produtivas. Esses cursos são concebidos como qualificações profissionais que se estruturam em matrizes curriculares articulando teoria e prática, buscando favorecer a permanência, o êxito e a inclusão social dos estudantes (IFRN, 2025).

Caracterizados como formações de curta duração, os Cursos FIC abrangem processos de aprendizagem e capacitação voltados tanto à formação inicial quanto à formação continuada. Eles possibilitam o desenvolvimento de conhecimentos profissionais, científicos e tecnológicos básicos, além da atualização e aperfeiçoamento de habilidades, contribuindo de maneira significativa para a inserção e reinserção de trabalhadores no mundo do trabalho e para a ampliação das oportunidades educacionais da comunidade (IFRN, 2025).

No Campus Canguaretama, os Cursos FIC simbolizam uma notável ação de extensão e inclusão, alcançando além de servidores e discentes o público externo, reforçando então o compromisso da Rede Federal com a democratização do conhecimento e de ampliação de oportunidades educacionais e profissionais.

Apesar da grande relevância destes cursos, observa-se que o processo de divulgação, inscrição e gerenciamento continua sendo realizado de forma manual e descentralizada, no qual as vagas são divulgadas apenas nos murais do campus ou por meio de formulários isolados (virtuais ou impressos), dificultando o acesso a comunidade interessada, reduzindo no número de inscritos, somado a falhas na comunicação institucional, comprometendo, deste modo, a efetividade das ações de inclusão e formação e reduzindo o alcance social dos cursos.

Com o intuito de superar essas limitações, torna-se necessário a busca por soluções tecnológicas que permitam informatizar e automatizar. O desenvolvimento de um sistema web voltado para o gerenciamento dos cursos, reunindo informações somente neste ambiente, com organização, transparência e acessibilidade, desde o processo de divulgação, inscrição e seleção, é uma alternativa altamente viável, já que essa ação facilitaria o acesso a comunidade interessada, tornaria a gestão mais eficiente, transparente e possibilitaria um maior alcance de formações. Além de que

o sistema proposto permitiria gerar relatórios e análises sobre os cursos, auxiliando a instituição a planejar melhor e aprimorar continuamente a oferta.

Deste modo, surgiu a ideia de criar um sistema on-line chamado IFic. Seu desenvolvimento tem como base três projetos de pesquisa: “Desenvolvimento de um Sistema Web para Gerenciamento de Cursos FICs do IFRN Campus Canguaretama”, projeto já concluído e aprovado no Edital nº 01/2022 – PROPI/RE/IFRN; “Sistema Web para Gerenciamento de Cursos FICs do IFRN Campus Canguaretama, etapa 2024”, projeto já concluído e aprovado no Edital nº 01/2024 – PROPI/RE/IFRN; e, por fim, “Desenvolvimento do iFIC: Site para Gerenciamento de Cursos FICs do IFRN, etapa 2025”, projeto em andamento e aprovado no Edital nº 01/2025 – PROPI/RE/IFRN.

Neste sentido, o presente estudo busca o desenvolvimento de um dashboard para o gerenciamento e análise dos cursos FIC, com foco inicial em nosso Campus Canguaretama. Ela permitirá visualizar dados quantitativos sobre os cursos já executados, possibilitando um acompanhamento mais sistemático e transparente das ações desenvolvidas. Sua base de informações para um dashboard será construída a partir da coleta e análise de dados relacionados aos cursos, de modo a subsidiar tanto a gestão acadêmica quanto a tomada de decisões institucionais.

A seguir, descreve-se a estrutura adotada neste estudo: o Capítulo 2 apresenta uma revisão da literatura sobre Análise de Dados; O capítulo 3 descreve a metodologia da utilizada na pesquisa; a discussão sobre os materiais e tecnologias utilizadas estão no Capítulo 4; o Capítulo 5 discute os resultados obtidos com o desenvolvimento do projeto; por fim, o Capítulo 6 apresenta as considerações finais da pesquisa.

1.1. OBJETIVO GERAL

O objetivo principal deste trabalho consiste em realizar uma análise exploratória e descritiva de dados sobre os cursos de Formação Inicial e Continuada (FIC) ofertados pelo IFRN, a partir das informações disponibilizadas pelo SUAP (Sistema Unificado de Administração Pública).

1.2. OBJETIVOS ESPECÍFICOS

Os objetivos específicos do presente trabalho são:

- Organizar e estruturar a base de dados dos cursos FIC do Campus Canguaretama definindo as variáveis necessárias para análise.
- Realizar o tratamento dos dados, incluindo padronização de campos, correção de inconsistências e tratamento de ausências (dados nulos), garantindo a qualidade da base.
- Aplicar análises descritivas e exploratórias para caracterizar a oferta de cursos, considerando recortes como período/ano, campus e demais dimensões disponíveis na base.
- Desenvolver e disponibilizar um dashboard para visualização e acompanhamento dos resultados, facilitando a interpretação dos indicadores analisados.

2. REFERENCIAL TEÓRICO

2.1. Análise de Dados

A análise de dados configura-se como um procedimento de caráter essencial e interdisciplinar, que compreende as etapas de coleta, tratamento, interpretação e representação das informações, com a finalidade de promover a extração de conhecimentos relevantes e a fundamentação de decisões orientadas por evidências empíricas (Silvestre, 2007). Além disso, observa-se a disponibilidade de uma ampla variedade de bibliotecas e *frameworks* computacionais que desempenham papel fundamental na otimização de processos analíticos. Esses recursos permitem a execução eficiente de cálculos matemáticos de elevada complexidade, a organização sistemática e o aprimoramento da qualidade dos dados, bem como a elaboração automatizada de representações gráficas e relatórios técnicos que auxiliam na interpretação dos resultados obtidos (Lopes, 2019).

A análise descritiva constitui-se como etapa fundamental na exploração inicial dos dados, ao organizar e sintetizar informações que, em seu estado bruto, não permitem inferências claras. Por meio dessa sistematização, torna-se possível delinear padrões gerais, identificar comportamentos recorrentes e reconhecer registros inconsistentes que possam comprometer etapas posteriores da investigação (Reis & Reis, 2002).

Essa fase utiliza-se de representações tabulares, gráficas e de medidas de síntese, que reduzem a complexidade do conjunto de dados sem prejuízo de sua essência informacional. Tais instrumentos favorecem a visualização estruturada das informações e ampliam a capacidade interpretativa do pesquisador, especialmente no reconhecimento de tendências e anomalias (Reis & Reis, 2002).

A seleção das técnicas descritivas depende da natureza da variável analisada. Variáveis qualitativas são usualmente representadas por distribuições de frequência e gráficos categóricos, enquanto variáveis quantitativas exigem representações como histogramas, ogivas, diagramas de pontos e estruturas de ramo-e-folhas, que evidenciam a distribuição e amplitude dos valores observados (Reis & Reis, 2002).

Três aspectos são centrais na caracterização de uma distribuição: tendência central, variabilidade e forma. A tendência central indica o valor mais representativo;

a variabilidade mensura o grau de dispersão dos dados; e a forma da distribuição revela sua simetria, assimetria e padrões de concentração, elementos essenciais para a compreensão do comportamento do fenômeno investigado (Reis & Reis, 2002).

Dessa maneira, a análise descritiva transcende o caráter introdutório e assume função estruturante na pesquisa, ao fornecer subsídios analíticos robustos que orientam a interpretação dos dados e fundamentam decisões metodológicas subsequentes (Reis & Reis, 2002).

A Análise Exploratória de Dados (AED) corresponde a uma etapa inicial e fundamental no estudo de um conjunto de informações, pois busca revelar aspectos que não são imediatamente perceptíveis. Seu objetivo é oferecer ao analista uma visão clara e intuitiva da estrutura dos dados, permitindo identificar características importantes de maneira rápida. Para isso, o uso de recursos visuais torna-se indispensável, já que a interpretação gráfica facilita o reconhecimento de padrões, contrastes, tendências, agrupamentos e possíveis associações, algo que seria mais difícil de perceber apenas por meio de valores numéricos (LOPES et al., 2019).

Além disso, a AED costuma ser apoiada por ferramentas computacionais, como a linguagem Python, que possibilitam explorar os dados de forma prática e eficiente. Essa exploração inicial contribui para compreender melhor o comportamento do conjunto analisado, localizar eventuais inconsistências e levantar hipóteses que poderão nortear análises posteriores (LOPES et al., 2019).

De modo geral, a AED pode ser entendida como um conjunto de técnicas voltadas à organização, descrição e interpretação de dados numéricos. Mesmo sendo procedimentos considerados simples, costumam apresentar grande robustez e boa resistência a situações com dados atípicos ou erros mais significativos, tornando-a especialmente útil em diferentes cenários analíticos (LOPES et al., 2019).

A linguagem Python tem se destacado como uma das principais ferramentas utilizadas em análise de dados, em grande parte por seu caráter de linguagem de propósito geral. Diferentemente de linguagens criadas especificamente para estatística, como o R, Python não se limita a aplicações analíticas, permitindo que seja empregada em diversas etapas do desenvolvimento de software. Essa versatilidade favorece sua adoção no ambiente corporativo, já que muitas equipes de tecnologia já dominam seus fundamentos, o que torna mais simples integrar

modelos analíticos aos sistemas existentes. Esse cenário contribui para que soluções de análise de dados desenvolvidas em Python sejam facilmente colocadas em produção, mantendo consistência entre o ambiente de desenvolvimento e o de execução (LOPES et al., 2019).

No contexto da análise de dados, Python se destaca pela ampla disponibilidade de bibliotecas especializadas, que oferecem recursos prontos para facilitar etapas como importação, tratamento e investigação dos dados (ANACONDA, 2023). Entre essas ferramentas, destacam-se Pandas e Matplotlib, utilizadas neste estudo por sua capacidade de organizar conjuntos de dados de forma eficiente e possibilitar a criação de visualizações que apoiam a interpretação dos resultados.

De acordo com a documentação do Pandas (2025), essa biblioteca de código aberto, distribuída sob licença BSD, oferece estruturas de dados eficientes e ferramentas práticas voltadas à análise de informações em Python. Ela permite armazenar grandes volumes de dados de maneira organizada e fornece diversos recursos para sua manipulação. Suas principais estruturas são Series e DataFrame, que possibilitam indexar, filtrar e consultar dados com rapidez. Além disso, o Pandas inclui funções para leitura e gravação em múltiplos formatos, tratamento de valores ausentes, combinação de tabelas, agregação e transformação de informações, entre outras operações. Por essas capacidades, tornou-se uma das bibliotecas mais utilizadas na área de ciência de dados, especialmente em etapas de limpeza, preparação e análise de conjuntos de dados.

Com base no que apresenta Hunter (2007), o Matplotlib surgiu como uma solução em Python voltada à produção de gráficos bidimensionais com nível adequado para aplicações, documentação e publicação científica. Seu funcionamento em diversos sistemas é favorecido pelo fato de ter sido projetado para uma linguagem já consolidada como multiplataforma.

Durante o desenvolvimento da biblioteca, Hunter tomou o Matlab como referência, já que utilizava essa ferramenta em atividades científicas e percebia limitações que poderiam ser superadas. A partir dessas observações, ele realizou um processo de engenharia reversa para criar um recurso em Python que mantivesse a qualidade gráfica e uma experiência de uso parecida com a do Matlab, mas que atendesse de forma mais ampla à comunidade que dependia desse tipo de visualização.

Além disso, o Matplotlib possui código aberto e é distribuído sob licença livre, o que facilita sua adoção. Para utilizá-lo, basta instalá-lo e importá-lo no ambiente Python, permitindo a criação imediata de visualizações sem barreiras de acesso (Hunter, 2007).

3. METODOLOGIA

O presente trabalho caracteriza-se como uma pesquisa aplicada, descritiva e exploratória. É aplicada porque tem como objetivo propor uma solução prática para um problema real: a falta de informatização no processo de divulgação e gerenciamento dos cursos de Formação Inicial e Continuada (FIC) do Campus Canguaretama do IFRN. É descritiva porque busca compreender e apresentar a forma como esses cursos são atualmente organizados e divulgados pela instituição. Por fim, é exploratória, pois investiga e analisa os dados disponíveis sobre os cursos, de modo a identificar informações relevantes que auxiliem na construção da dashboard proposta.

A pesquisa apresenta uma abordagem quantitativa, pois trabalha com dados institucionais estruturados em planilha, contendo informações numéricas e categóricas sobre os cursos FIC. Foi também realizada uma pesquisa bibliográfica com o intuito de embasar teoricamente o trabalho, buscando compreender a relevância dos cursos FIC e o papel das tecnologias na gestão educacional. Essa pesquisa foi conduzida em fontes como o Google Scholar e o repositório institucional do IFRN, utilizando termos relacionados à temática, como “Formação Inicial e Continuada”, “educação profissional”, “inclusão digital” e “sistemas educacionais informatizados”. Além disso, o orientador do trabalho disponibilizou uma planilha institucional contendo os dados utilizados na análise.

Em seguida, os dados dessa planilha foram organizados e ajustados no ambiente Google Colab, de forma a corrigir possíveis inconsistências e padronizar o formato das colunas. Essa etapa foi essencial para garantir que as informações pudessem ser analisadas corretamente e visualizadas de maneira clara.

Por fim, foi realizada uma análise exploratória dos dados, com o objetivo de compreender o perfil dos cursos FIC ofertados pelo campus, identificar padrões, número de vagas, períodos de funcionamento e outras características relevantes. A partir dessa análise, foram geradas visualizações e informações que servirão de base para a construção da dashboard, contribuindo para uma gestão mais transparente e eficiente dos cursos FIC.

4. RESULTADOS

4.1. *Dataset* analisado

Os conjuntos de dados (*dataset*) analisados neste trabalho foram obtidos no dia 04 de fevereiro de 2026 a partir do Sistema Unificado de Administração Pública (SUAP), o sistema institucional utilizado pelo IFRN para o gerenciamento de informações acadêmicas, administrativas e educacionais, concentrando dados oficiais referentes aos cursos servidores e estudantes da instituição, por intermédio da Secretaria Acadêmica do IFRN - Campus Canguaretama.

O material coletado corresponde aos cursos de Formação Inicial e Continuada (FIC) ofertados pelo IFRN no período de 2020 a 2026. Os dados foram exportados em formato XLS, por meio do módulo de Relatórios de Ensino do SUAP. Devido ao tempo de processamento do SUAP e garantir dados mais precisos, optou-se por exportar as listas de matrículas em cursos FIC por ano.

As informações contidas no conjunto de dados abrangem aspectos acadêmicos e operacionais dos cursos FIC, permitindo uma visão ampla sobre sua oferta, organização e características gerais. Ao total, os sete *datasets* totalizam 79.231 linhas e 17 colunas, descritas a seguir.

- Col1 - Matrícula (dado anonimizado pelo próprio SUAP);
- Col2 - Nome (dado anonimizado pelo próprio SUAP);
- Col3 - Ano de Conclusão;
- Col4 - Ano de Ingresso;
- Col5 - Campus;
- Col6 - Código curso;
- Col7 - Data de Conclusão;
- Col8 - Data de Integralização
- Col9 - Descrição do Curso;
- Col10 - Diretoria;
- Col11 - Edital de Ingresso;
- Col12 - Matriz;
- Col13 - Modalidade;
- Col14 - Período Letivo de Integralização;

- Col15 - Situação no Curso;
- Col16 - Turno;

A Tabela 1 apresenta a quantidade de linhas por dataset exportado.

Tabela 1. Quantidade de linhas nos datasets por ano

Ano	Linhas
2020	14203
2021	5130
2022	5610
2023	9298
2024	12395
2025	32245
2026	350

Fonte: Autoria própria (2026)

4.2. Preparação dos Dados

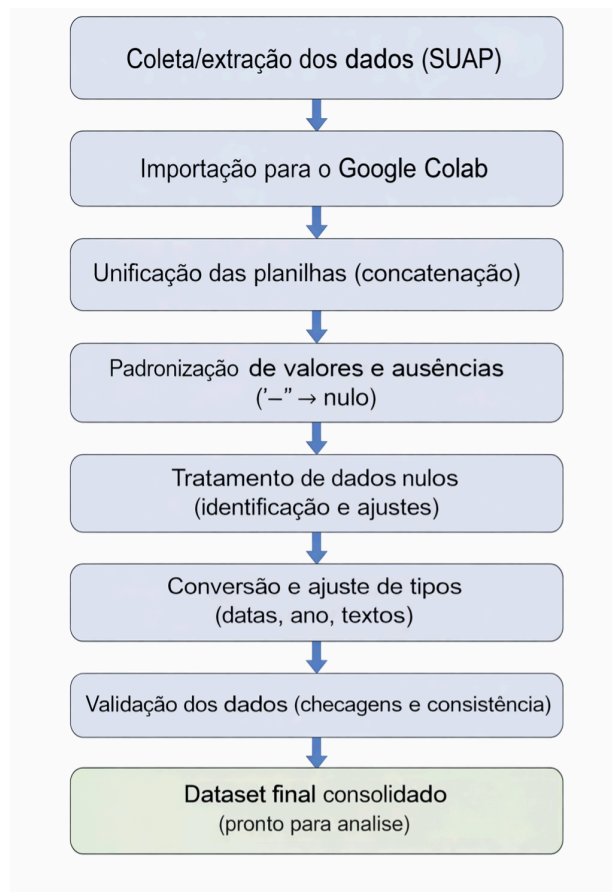
Antes da realização da análise, os dados extraídos do SUAP precisaram ser preparados, com o objetivo de garantir a qualidade, a consistência e a confiabilidade das informações. Embora o *dataset* estivesse estruturado em formato de planilha eletrônica, foi necessário realizar ajustes prévios, como inspeção da estrutura dos dados, verificação de valores ausentes e atípicos, padronização de caracteres e conversão dos tipos das variáveis.

A preparação dos dados foi realizada no ambiente Google Colab¹ (ANEXO 01), utilizando a linguagem Python e as bibliotecas Pandas, NumPy, Matplotlib e Seaborn, que possibilitaram a organização, o tratamento e a visualização das informações de forma eficiente, servindo como base para as análises exploratórias apresentadas nas seções seguintes.

¹ Google Colab (Google Colaboratory). Disponível em: <https://colab.research.google.com/>. Acesso em: 04 mar. 2026.

Para tornar mais clara a sequência adotada na preparação da base, a Figura 01 apresenta, de forma ilustrativa, as principais etapas executadas até a obtenção do dataset consolidado e pronto para análise.

Figura 01. Etapas de preparação de dados



Fonte: Autoria própria (2026)

Após essas etapas, a base consolidada foi utilizada nas análises descritivas e exploratórias apresentadas nas seções seguintes.

4.2.1. Concatenação dos Datasets

Durante a etapa de coleta dos dados, foram geradas sete planilhas distintas, cada uma correspondente a um ano do período analisado, compreendido entre 2020 e 2026. Essa organização anual facilitou a extração dos dados a partir do sistema, porém, para a realização de uma análise conjunta e comparativa ao longo do tempo, tornou-se necessário reunir todas essas informações em um único conjunto de dados.

Dessa forma, as sete planilhas foram concatenadas, resultando em um *dataset* único consolidado, contendo todos os registros dos cursos de Formação Inicial e Continuada (FIC) ofertados no período analisado. A concatenação permitiu uma visão global dos dados, possibilitando a aplicação das técnicas de análise descritiva e exploratória de forma mais consistente e abrangente.

Após a junção das planilhas, foi realizado o *reset* dos índices das linhas, com o objetivo de reorganizar a numeração dos registros. Esse procedimento foi necessário para assegurar que o conjunto de dados final representasse corretamente todo o período analisado e evitar problemas relacionados à duplicidade ou à descontinuidade dos índices, garantindo a correta contagem dos dados e a integridade das análises estatísticas realizadas posteriormente.

4.2.2. Substituição de Caracteres textuais

Durante a verificação inicial do conjunto de dados, foi identificado que algumas colunas apresentavam o caractere textual “-” (hífen) como forma de indicar a ausência de informação. Esse tipo de representação, embora comum em registros administrativos, pode causar inconsistências durante o processo de análise, especialmente em colunas que deveriam conter valores numéricos ou datas.

A presença desse caractere inviabiliza a conversão adequada dos dados e pode comprometer a aplicação de técnicas estatísticas e a geração de visualizações, como gráficos do tipo boxplot, que exigem valores numéricos válidos para seu correto funcionamento.

Diante deste cenário, optou-se por substituir o caractere “-” pelo tipo nulo, padronizando o conjunto de dados e permitindo a correta conversão das colunas para os tipos apropriados. Essa etapa foi essencial para garantir a integridade das análises realizadas, possibilitando a identificação de valores atípicos (*outliers*) e assegurando que os resultados obtidos representassem de forma mais fiel o comportamento real dos dados analisados.

4.2.3. Dados Nulos

Inicialmente, foi realizada uma busca por dados nulos no conjunto de dados, tanto de forma geral quanto por coluna, com o objetivo de identificar possíveis lacunas e inconsistências nas informações disponibilizadas. Essa verificação permitiu mapear quais atributos apresentavam valores ausentes e em qual quantidade eles ocorriam ao longo do *dataset*. A Tabela 2 demonstra quais colunas, a quantidade e a proporção de dados nulos localizados.

Tabela 2. Quantidade de Dados Nulos

Coluna	Quantidade	% em relação ao total
ANO DE CONCLUSÃO	57483	72,6%
DATA DE CONCLUSÃO DE CURSO	12605	15,9%
DATA DE INTEGRALIZAÇÃO	79231	100%
EDITAL DE INGRESSO	53695	67,8%
MATRIZ	44707	56,4%
PERÍODO LETIVO DE INTEGRALIZAÇÃO	79231	100%
TURNOS	26694	33,7%

Fonte: Autoria Própria (2026)

É possível verificar que determinadas colunas apresentam uma quantidade elevada de registros sem preenchimento, o que evidencia que nem todas as informações são registradas de forma padronizada ao longo de todo o período analisado. Esse comportamento é comum em bases de dados institucionais amplas e em constante crescimento, especialmente quando envolvem cursos ofertados em diferentes anos e sob distintos contextos administrativos.

Destaca-se que as colunas Data de Integralização e Período Letivo de Integralização concentram o maior número de valores nulos, abrangendo praticamente a totalidade dos registros. Nesse contexto, a ausência dessas informações sugere que esses campos não são usualmente registrados ou não se mostram relevantes para cursos do tipo FIC, sendo mais característicos de outras

modalidades ofertadas pela instituição que possuem acompanhamento formal de integralização acadêmica.

Observa-se também uma quantidade significativa de dados nulos nas colunas Ano de Conclusão e Edital de Ingresso. Em relação ao Ano de Conclusão, a ausência pode estar associada, por um lado, a ofertas que não foram finalizadas ou não tiveram a conclusão registrada no sistema; por outro, também foram encontrados registros de cursos mais antigos que, embora possivelmente já encerrados, permanecem sem esse preenchimento. Isso indica inconsistências no registro histórico e reforça que o “Ano de Conclusão” seria uma variável importante para análises mais estratégicas, como acompanhamento de finalização e avaliação de efetividade das ofertas.

Já quanto ao Edital de Ingresso, uma possível explicação para a grande quantidade de nulos está relacionada ao fato de que o cadastro de cursos FIC no SUAP passou a ser adotado de forma mais recente, enquanto em períodos anteriores o controle de editais e ofertas podia ocorrer por outros meios, dificultando o levantamento retroativo e a atualização completa dessas informações no sistema.

As colunas Matriz e Turno, por sua vez, apresentam um volume intermediário de valores ausentes, indicando que tais informações estão disponíveis apenas para parte dos cursos. Essa variação pode estar associada a mudanças nos padrões de cadastro ao longo dos anos, além de particularidades operacionais de determinadas ofertas.

Apesar disso, a variável Turno foi utilizada nas análises para permitir a avaliação da distribuição das ofertas por turno ao longo do tempo, considerando apenas os casos em que o preenchimento estava disponível. Ela também foi mantida no dataset por possuir relevância institucional e potencial de uso em análises futuras.

A análise dos dados nulos, portanto, contribui para uma compreensão mais adequada das limitações do *dataset* e orientou as decisões adotadas nas etapas seguintes do estudo, reforçando a importância da preparação e avaliação da qualidade dos dados para a obtenção de resultados mais consistentes e confiáveis.

4.2.4. Conversão de Tipos de Dados

Durante a etapa de preparação, verificou-se a necessidade de realizar a conversão dos tipos de algumas colunas, uma vez que determinadas informações estavam registradas em formato textual, o que poderia comprometer as análises estatísticas e temporais propostas neste estudo.

A coluna ‘Ano de Ingresso’ foi convertida para o tipo numérico, com o objetivo de possibilitar a aplicação de técnicas estatísticas, como identificação de valores atípicos (*outliers*) e a análise da distribuição dos cursos ao longo dos anos. Essa conversão foi essencial para garantir que os valores pudessem ser corretamente interpretados e comparados, evitando inconsistências decorrentes de registros não numéricos ou inválidos.

De forma semelhante, a coluna ‘Data de Conclusão de Curso’ foi convertida para o formato de data. Essa transformação teve como finalidade padronizar a variável e garantir sua correta interpretação no conjunto de dados. Embora não tenham sido realizados cálculos ou análises específicas com base nessa coluna no escopo deste trabalho, a conversão foi mantida por se tratar de uma informação relevante para análises futuras, especialmente aquelas que envolvam comparações temporais e avaliação de períodos de conclusão, quando houver preenchimento adequado do campo.

A conversão adequada dos tipos de dados mostrou-se, portanto, uma etapa fundamental para assegurar a consistência do conjunto de dados e a confiabilidade das análises exploratórias realizadas. Ao padronizar os formatos numéricos e temporais, foi possível garantir que os resultados obtidos refletissem de maneira mais fiel o comportamento real dos cursos analisados.

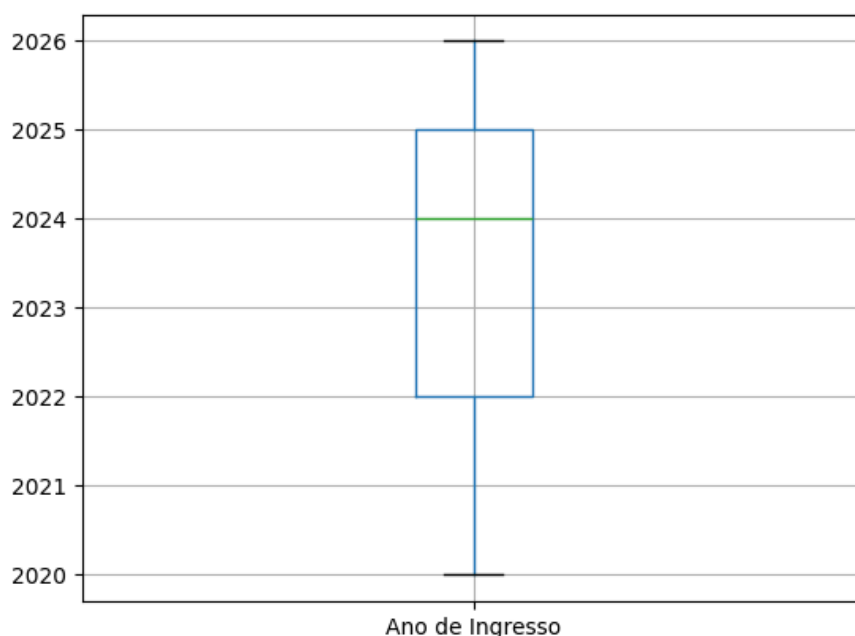
4.2.5. Dados atípicos (*outliers*)

Posteriormente, foi realizada uma verificação de valores atípicos (*outliers*) nas variáveis numéricas do conjunto de dados, com o objetivo de identificar possíveis registros fora do padrão que pudessem comprometer a interpretação dos resultados estatísticos e das visualizações. Considerando a natureza da base utilizada, a avaliação foi direcionada à coluna “Ano de Ingresso”, por se tratar de um

campo numérico diretamente associado à dimensão temporal da análise e por ser amplamente utilizado nas segmentações por ano.

Para essa etapa, foi empregado o diagrama de caixa (*boxplot*), por se tratar de uma técnica exploratória adequada para inspeção de dispersão e detecção de valores extremos. Nesse tipo de gráfico, valores potencialmente atípicos costumam se manifestar como pontos isolados (“bolinhas”) além dos limites, que representam o intervalo esperado de variação dos dados (com base nos quartis). A Figura 2. apresenta o resultado da análise no gráfico.

Figura 2. Resultado da análise no gráfico



Fonte: Autoria própria (2026)

O boxplot da variável “Ano de Ingresso” indica que os registros estão concentrados em uma faixa central de valores, porém com assimetria na distribuição. Observa-se que a mediana está posicionada acima do centro da caixa, mais próxima do terceiro quartil (Q3), o que sugere maior concentração dos dados nos anos mais recentes (parte superior da caixa). Em contrapartida, nota-se maior extensão na parte inferior, indicando uma cauda para anos anteriores, com menor frequência e maior dispersão. Como não aparecem pontos isolados fora dos limites, conclui-se que não há valores discrepantes (outliers) visíveis para essa variável no período analisado. Dessa forma, os dados do “Ano de Ingresso” foram considerados

consistentes, não sendo necessário aplicar tratamento adicional, como remoção, correção ou truncamento de registros.

4.3. Análises Descritivas e Exploratórias

4.3.1. Modalidades

Para identificar os tipos de cursos presentes na base de dados, foi inicialmente verificada a coluna “Modalidade”. Observou-se que essa coluna apresenta apenas o valor “FIC” em todos os registros, não permitindo, por si só, distinguir diferentes categorias de oferta. Dessa forma, a diferenciação entre os tipos de cursos precisou ser realizada a partir da coluna “Descrição do Curso”, uma vez que seus primeiros caracteres e expressões específicas indicam o formato/programa do curso.

A partir da inspeção dos valores únicos dessa coluna e, posteriormente, de consulta à Pró-Reitoria de Ensino do IFRN, foi possível identificar padrões de nomenclatura associados a diferentes tipos de cursos e respectivas cargas horárias. Foram reconhecidos, ao todo, seis grupos principais:

- “FIC-”: cursos com carga horária inferior a 160 horas;
- “FIC+” e “FIC +”: cursos com carga horária igual ou superior a 160 horas;
- “ProITEC”: cursos do Programa de Iniciação Tecnológica e Cidadania, com 165 horas;

Cursos identificados nominalmente na descrição, como:

- “FIC em Produtor(a) de Derivados do Leite [Mulheres Mil 2024]”: curso com 172 horas;
- “FIC em Artesã em Bordado à Mão [Mulheres Mil 2018]” e “FIC em Artesã em Bordado à Mão [Mulheres Mil 2019]”: curso com 214 horas;
- “Curso de Formação Inicial e Continuada do Programa Nacional de Promoção de Igualdade de Oportunidade para o acesso de estudantes da rede pública de ensino à Rede Federal de Educação Profissional, Científica e Tecnológica – PartiuLF” (descrição extensa do programa): curso com 320 horas.

Apesar dessa diversidade de denominações, para fins desta análise descritiva e exploratória, optou-se por consolidar os cursos em duas categorias comparáveis, com base na carga horária:

- FIC-: cursos com menos de 160 horas;
- FIC+: cursos com 160 horas ou mais.

Essa padronização foi adotada para simplificar a interpretação dos resultados e viabilizar comparações mais consistentes ao longo do tempo, entre campi e entre turnos, mantendo a rastreabilidade dos cursos por meio da coluna “Descrição do Curso”.

4.3.2. Informações Gerais

Após a padronização dos cursos em duas categorias, foi realizada uma síntese quantitativa para compreender a distribuição geral da oferta no período analisado.

De forma agregada, identificou-se um total de 408 cursos FIC- e 77 cursos FIC+. Esses valores indicam predominância de cursos com menor carga horária, o que sugere que, no recorte observado, a oferta se concentrou majoritariamente em formações de curta duração, enquanto as formações com carga horária igual ou superior a 160 horas aparecem em menor volume. A Tabela 3 apresenta os números de cursos que foram ofertados por ano.

Tabela 3 . Quantidade de cursos ofertados por ano conforme a categoria (FIC- e FIC+)

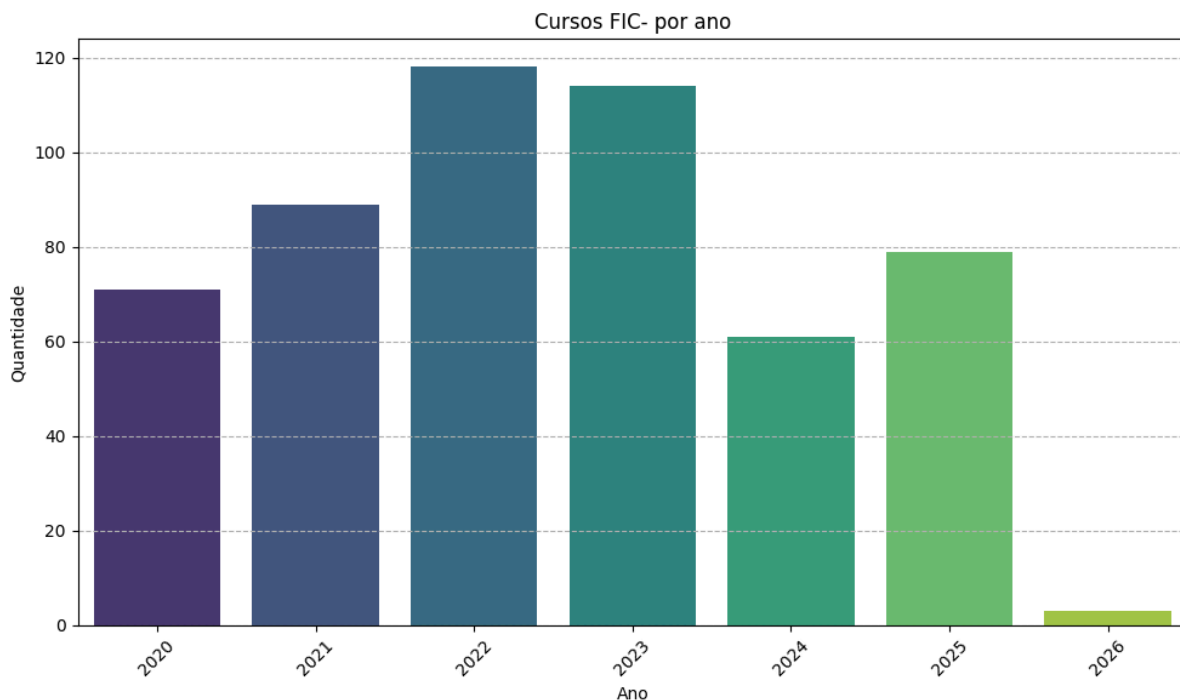
Ano	FIC-	FIC+
2020	71	17
2021	89	13
2022	118	17

2023	114	17
2024	61	33
2025	79	34
2026	3	1

Fonte: Autoria própria(2026)

Ao observar a distribuição por ano de ingresso, nota-se que os cursos FIC- apresentam maior variação ao longo do período. Em 2020 foram registrados 71 cursos FIC-, crescendo em 2021 (89) e atingindo o maior valor em 2022 (118). Em 2023 o volume permanece elevado (114), mas em 2024 ocorre uma redução expressiva (61). Em 2025 há recuperação parcial (79), enquanto 2026 apresenta um número reduzido (3), o que pode estar associado ao fato de se tratar de um ano ainda em andamento ou com registros incompletos no momento da extração dos dados. A Figura 3 demonstra essas informações através de um gráfico de barras.

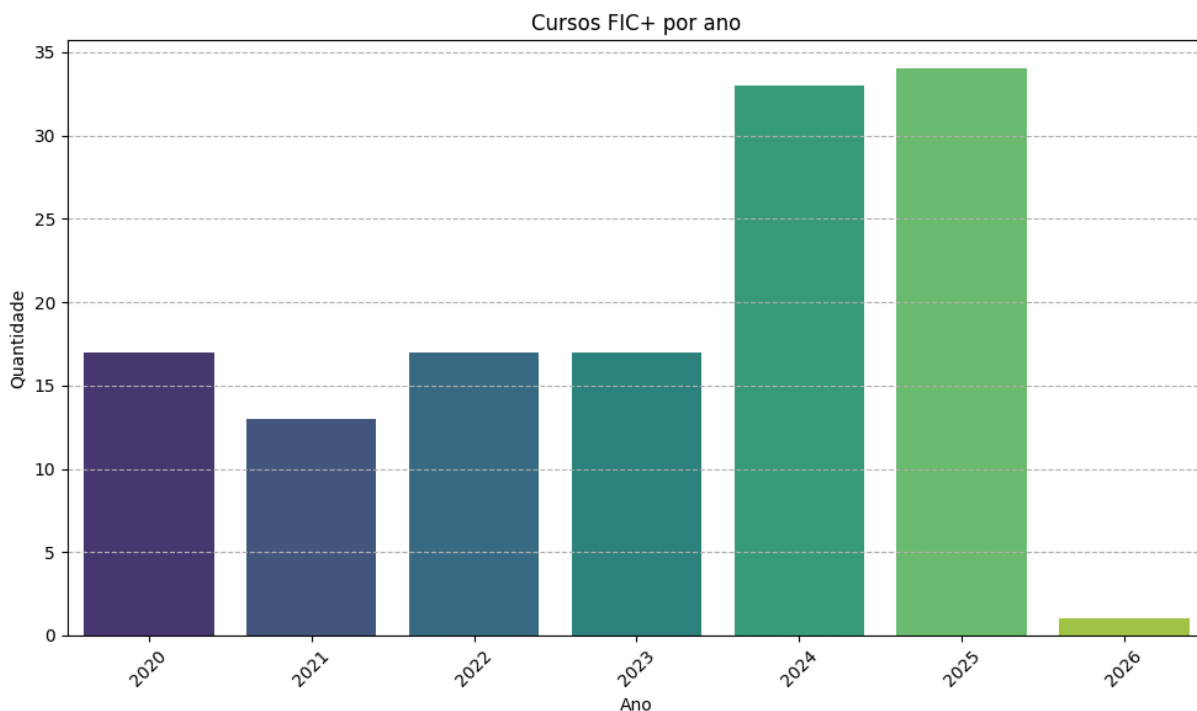
Figura 3. Cursos FIC com carga horária inferior a 160 horas por ano



Fonte: Autoria própria (2026)

Quanto aos cursos FIC+, observa-se um comportamento mais contido nos primeiros anos do recorte: 2020 (17), 2021 (13), 2022 (17) e 2023 (17). A partir de 2024, há um aumento relevante, com 33 cursos FIC+ e, em 2025, 34 cursos, sugerindo maior presença de cursos com carga horária igual ou superior a 160 horas nos anos mais recentes da série. Em 2026, aparece apenas 1 curso FIC+, novamente com a ressalva de possível parcialidade do ano. A Figura 4 possui um gráfico de barras com essas informações.

Figura 4. Cursos FIC com carga horária igual ou superior a 160 horas por ano



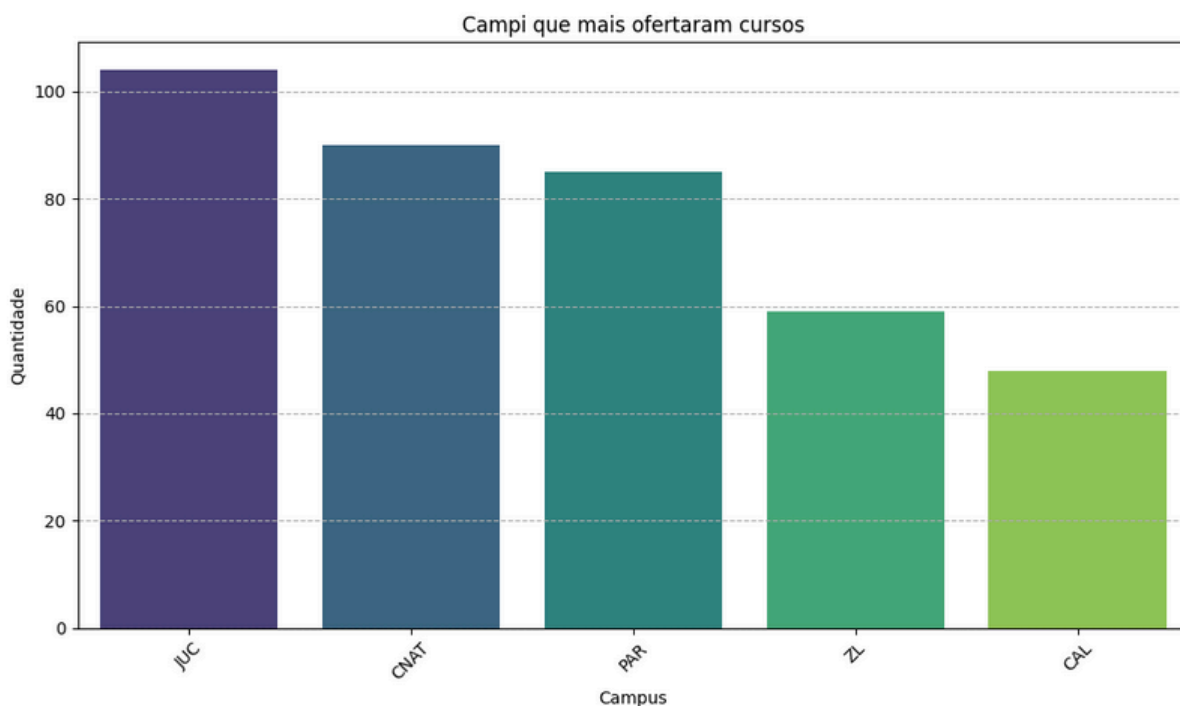
Fonte: Autoria própria (2026)

Em síntese, os dados indicam que a oferta anual de cursos, ao longo do período analisado, é fortemente marcada pela predominância dos cursos FIC-, cenário que se mantém de forma mais evidente até 2023. As figuras reforçam esse comportamento histórico, mostrando que, por vários anos, a maior parte das ofertas esteve concentrada em cursos com menor carga horária. Entretanto, a partir de 2024, observa-se um crescimento mais expressivo dos cursos FIC+, evidenciando uma intensificação recente desse tipo de oferta, o que pode estar relacionado a mudanças na estratégia institucional, à implementação de novos programas, ou ainda a ajustes na organização e na carga horária dos cursos ofertados. Esse contraste entre a estabilidade histórica do FIC- e o avanço recente do FIC+ sugere uma reconfiguração no perfil de oferta nos anos mais recentes. Assim, esses achados servem como base para aprofundar as análises nas subseções seguintes, explorando recortes por campus, turno e outros critérios, com o objetivo de verificar se essa mudança ocorre de maneira generalizada em toda a instituição ou se está concentrada em unidades e contextos específicos.

4.3.3. Ofertas

Com o objetivo de identificar quais unidades apresentaram maior volume de oferta no período analisado (2020–2026), os dados foram agrupados por Campus e, em seguida, foi contabilizada a quantidade de cursos únicos (com base na coluna “Descrição do Curso”). Esse procedimento permite observar, de forma mais objetiva, quais campi concentraram a maior diversidade de cursos ofertados no recorte temporal. A Figura 5 apresenta os cinco campi que mais ofertaram cursos no período de 2020 a 2026, considerando a contagem de cursos únicos com base na coluna “Descrição do Curso”.

Figura 5. Cinco campi que mais ofertaram curso de 2020 a 2026



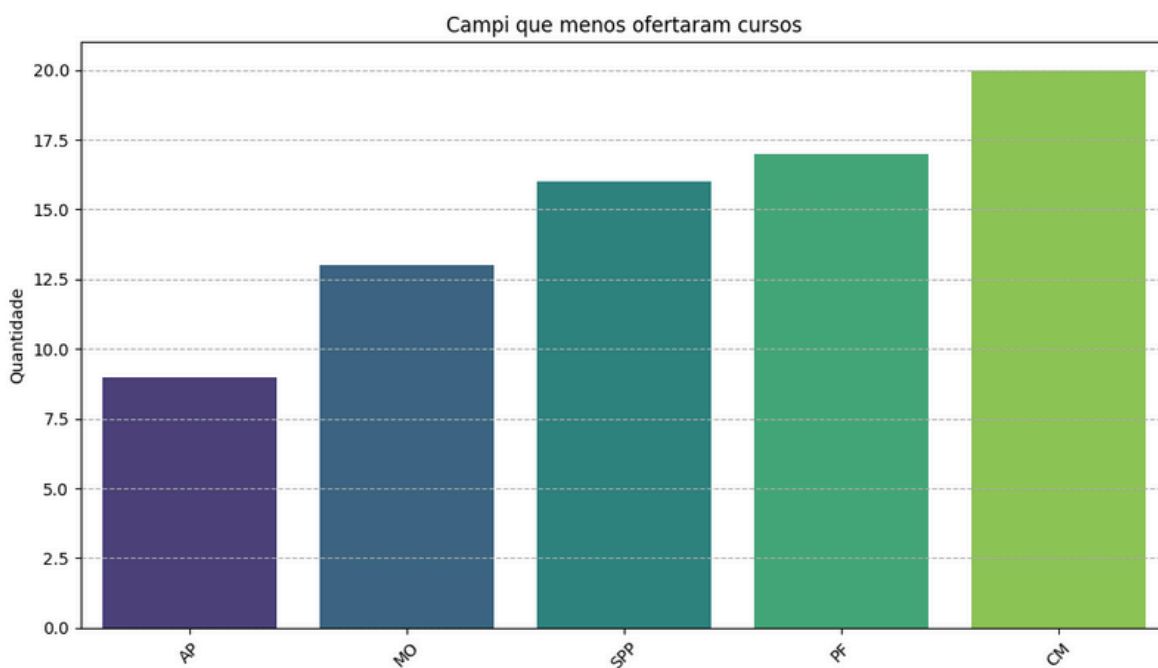
Fonte: Autoria própria(2026)

Os resultados indicam que os cinco campi com maior número de ofertas foram: “Jucurutu” (104), “Natal Central” (90), “Parnamirim” (85), “Zona Leste” (59) e “Cidade Alta” (48). Esse ranking evidencia uma concentração da oferta em poucas unidades, com destaque para Jucurutu, que apresenta o maior quantitativo no período, seguido por “Natal Central” e “Parnamirim” que aparecem em patamares próximos.

De modo geral, essa distribuição sugere que parte relevante das ofertas do período se encontra centralizada nesses campi, o que reforça a importância de, nas análises subsequentes, observar se essa concentração está associada a fatores como estrutura institucional, demanda local, programas específicos ou perfil de cursos ofertados.

Para complementar a análise de distribuição das ofertas, também foi identificado o grupo de campi com menor número de cursos ofertados no período de 2020 a 2026. Assim como na etapa anterior, a contagem foi feita a partir da quantidade de cursos únicos (com base na coluna “Descrição do Curso”), o que permite observar a diversidade de ofertas por campus, e não o volume de matrículas. A Figura 6 apresenta os cinco campi com menor número de cursos únicos ofertados no período analisado.

Figura 6. Cinco campi com menor número de cursos ofertados



Fonte: Autoria própria(2026)

Os resultados apontam que os cinco campi com menor quantitativo de ofertas foram: “Apodi” (9), “Mossoró” (13), “São Paulo do Potengi” (16), “Pau dos Ferros” (17) e “Ceará-Mirim” (20). Quando comparados ao grupo dos campi com maior oferta, esses valores evidenciam uma diferença considerável na distribuição

de cursos ao longo do período analisado, sugerindo que a oferta não ocorreu de forma homogênea entre as unidades.

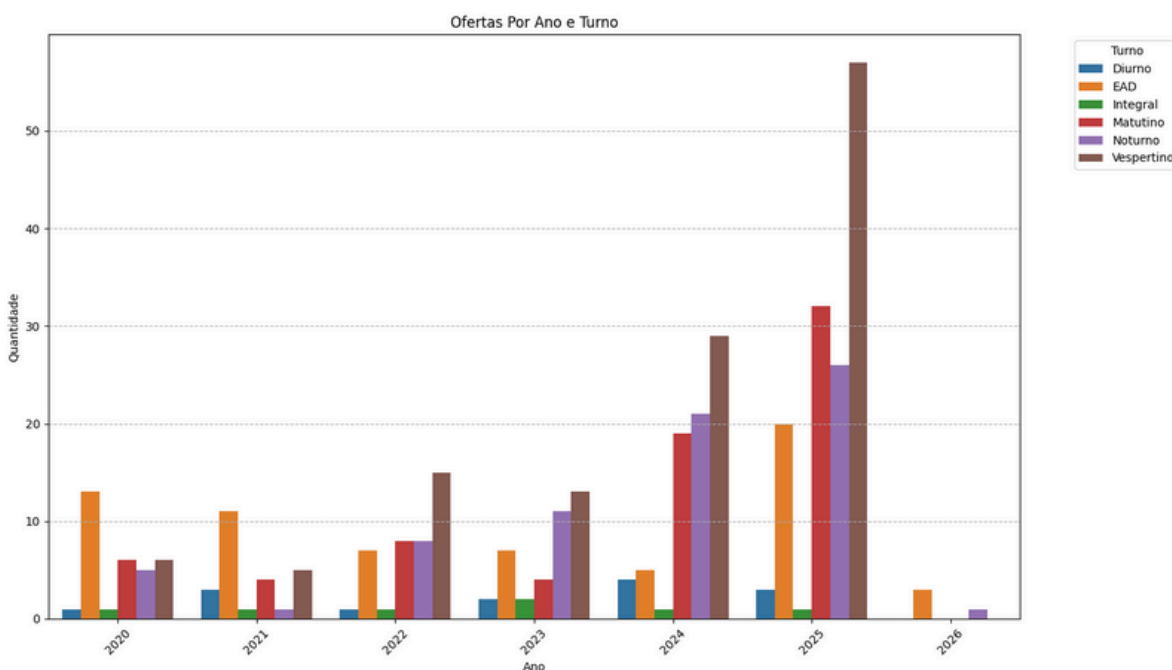
4.3.4. Turnos

A coluna “Turno” apresenta 26.694 dados nulos. Mesmo com essa limitação, optou-se por manter a variável na análise, pois ela ainda contém um volume relevante de registros preenchidos, permitindo observar padrões de oferta quando o turno está informado.

Dessa forma, foi realizado o cálculo do número de cursos ofertados por ano considerando os turnos, contabilizando-se os cursos únicos (com base na coluna “Descrição do Curso”) em cada combinação de Ano de Ingresso e Turno.

O gráfico resultante está na Figura 7 e apresenta a distribuição anual das ofertas por turno.

Figura 7. Distribuição anual das ofertas por turno



Fonte: Autoria própria (2026)

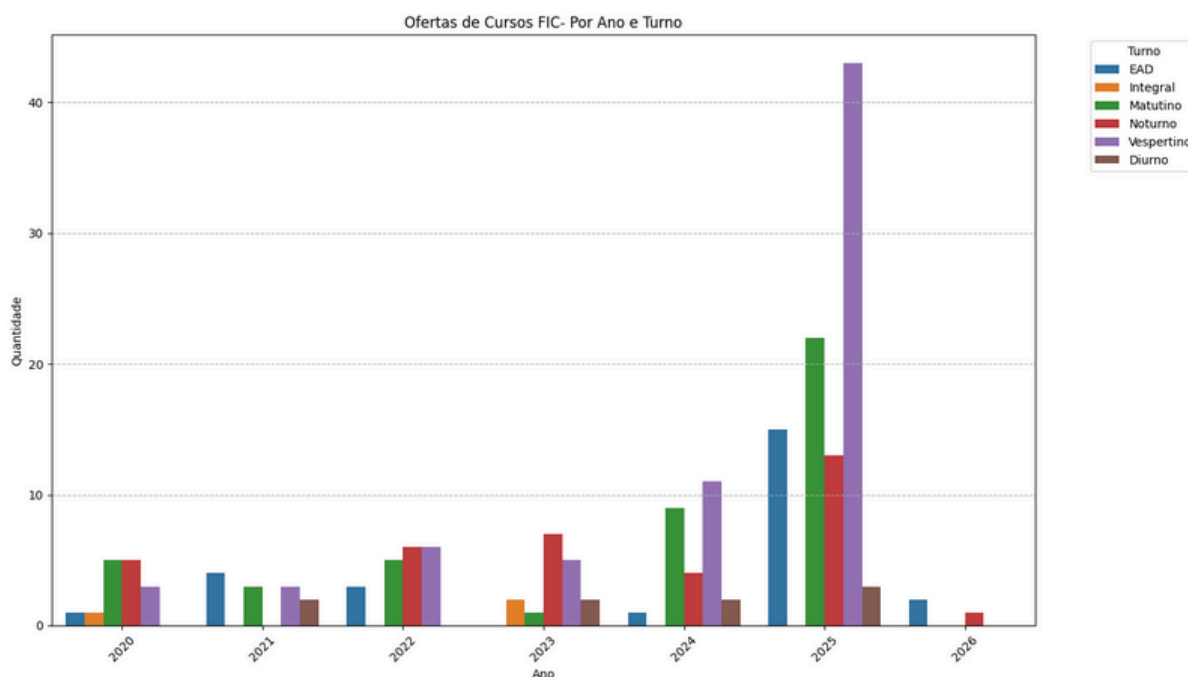
A Figura 6 sintetiza a evolução da oferta de cursos por turno entre 2020 e 2026. Observa-se que o turno vespertino se destaca como o principal período de

realização dos cursos ao longo do recorte, com crescimento mais evidente nos anos recentes - especialmente em 2024 e, de forma mais acentuada, em 2025, quando ocorre o maior pico de diversidade de cursos nesse turno.

Em contraste, os demais turnos apresentam volumes menores e, em geral, mais distribuídos ao longo do tempo. Esse comportamento sugere que a oferta não se distribui de forma homogênea entre os turnos, indicando uma concentração institucional (ou operacional) mais forte no período da tarde.

A Figura 8 apresenta a distribuição das ofertas de cursos FIC- por ano de ingresso e turno, considerando apenas os registros com essa informação preenchida.

Figura 8. Oferta de cursos FIC com carga horária inferior a 160 hrs por ano e turno



Fonte: Autoria própria (2026)

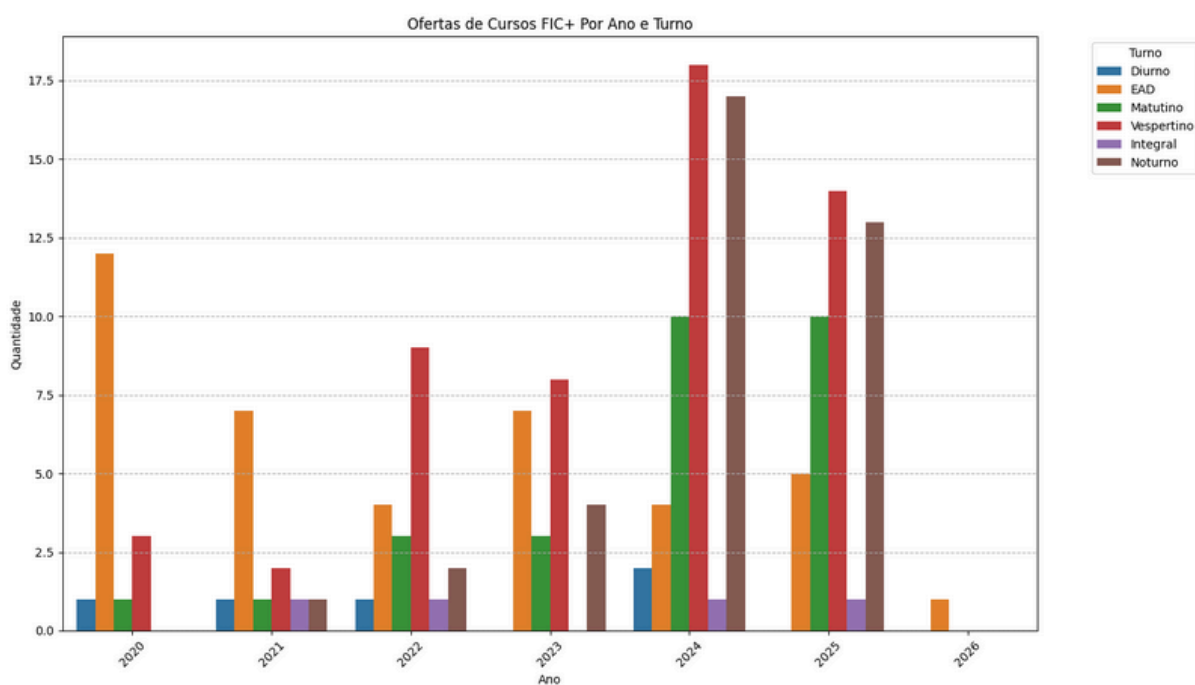
Ao restringir a análise aos cursos FIC- (carga horária inferior a 160 horas), a Figura 7 mantém o padrão observado anteriormente, mas com um destaque ainda mais claro: o vespertino concentra a maior diversidade de ofertas, com forte elevação em 2025.

Os turnos matutino e noturno também apresentam crescimento em

2024–2025, porém em patamares inferiores ao vespertino. Em síntese, entre os cursos de menor carga horária, a oferta tende a se organizar majoritariamente no período da tarde, com intensificação mais visível nos anos finais do recorte.

A distribuição das ofertas de cursos FIC+ por ano de ingresso e turno é apresentada na Figura 9.

Figura 9. Cursos ofertados com carga horária igual ou superior a 160 horas por turno e ano



Fonte: Autoria própria(2026)

A Figura 8 apresenta os cursos FIC+ (carga horária igual ou superior a 160 horas) e revela uma dinâmica parcialmente distinta. Diferentemente do grupo FIC-, aqui a distribuição entre turnos se mostra menos concentrada exclusivamente no vespertino, sobretudo a partir de 2024, quando outros turnos passam a ganhar relevância na diversidade de cursos ofertados.

Ainda assim, o vespertino permanece entre os turnos com maior presença, especialmente nos anos de maior expansão do conjunto analisado. Isso sugere que a predominância do período da tarde não se limita aos cursos de menor

duração, mas tende a se manter também entre ofertas mais longas, ainda que, no FIC+, a distribuição aparente ser relativamente mais equilibrada em alguns anos.

4.3.5. Ofertas de Cursos FIC no Campus Canguaretama

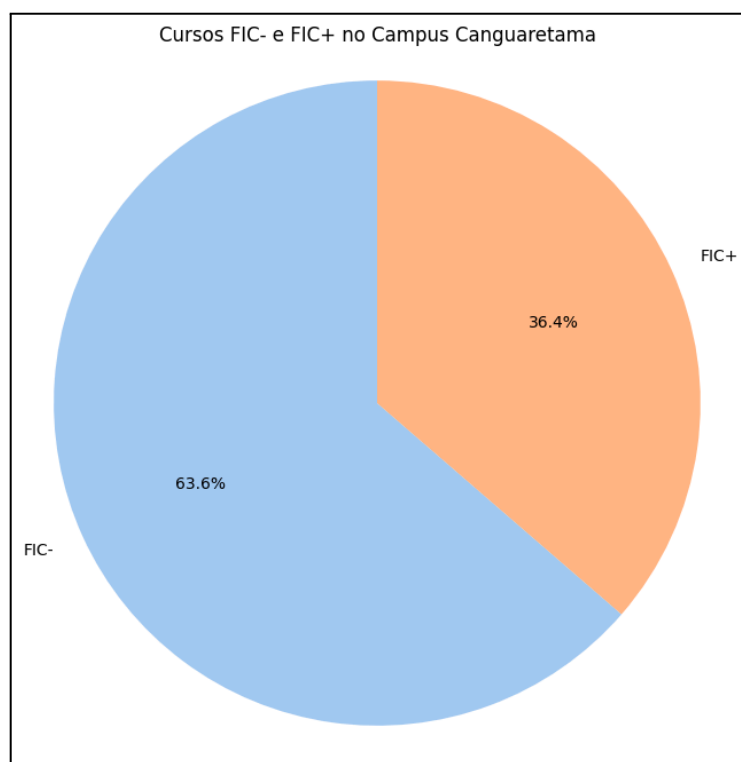
Com o objetivo de aprofundar a análise em nível local, foi realizado um recorte específico para o Campus Canguaretama (CANG). Nesse recorte, identificou-se um total de 22 cursos distintos no período analisado, considerando a contagem de cursos únicos a partir da coluna “Descrição do Curso”. Essa contagem representa a diversidade de ofertas registradas na base para o campus, e não o número de matrículas ou participantes.

A listagem de cursos evidencia que a oferta do campus é composta por formações de diferentes perfis e programas, incluindo cursos de curta duração (FIC-), cursos com carga horária mais elevada (FIC+), além de iniciativas específicas, como ProI TEC, Mulheres Mil, PRONATEC e PartiuIF. Entre os cursos identificados, destacam-se, por exemplo, formações introdutórias e de base, como Introdução ao Letramento Digital, Matemática Básica, LIBRAS para Incluir e Língua Inglesa (nível elementary), bem como cursos relacionados a tecnologia e inovação, como Desenvolvimento de Web e Mobile e atividades de robótica educacional. Também aparecem ofertas com foco em qualificação profissional e programas sociais, como Camareira em Meios de Hospedagem (Mulheres Mil 2024), Agricultor Familiar (Mulheres Mil), Agente de Desenvolvimento Cooperativista (Mulheres Mil 2024) e Eletricista de Sistemas de Energias Renováveis (PRONATEC 2023), além do curso do programa PartiuIF, que possui maior carga horária e caráter estruturante.

Ao consolidar os cursos por categoria de carga horária (conforme a padronização adotada anteriormente), observou-se que o campus apresenta predominância de cursos FIC-, com 14 ofertas (aproximadamente dois terços do total), enquanto os cursos FIC+ somam 8 ofertas. Esse resultado sugere que, no recorte analisado, a dinâmica de oferta do campus é majoritariamente composta por cursos com menor carga horária, embora também haja presença relevante de cursos mais longos, associados principalmente a programas específicos.

A Figura 10 apresenta a distribuição dos 22 cursos ofertados pelo Campus Canguaretama entre as categorias FIC- e FIC+.

Figura 10. Distribuição dos 22 cursos do Campus Canguaretama entre as categorias FIC- e FIC+.



Fonte: Autoria própria(2026)

O gráfico de pizza apresentado para o Campus Canguaretama tem como objetivo mostrar, de forma rápida e visual, como os 22 cursos do campus se dividem entre as categorias FIC- e FIC+ (de acordo com a padronização adotada na análise).

Nesse recorte, foram identificados 14 cursos classificados como FIC- e 8 cursos classificados como FIC+. Ao transformar esses valores em proporção, o gráfico evidencia que a maior parte das ofertas do campus está concentrada em cursos FIC-, que representam 63,6% do total (14 de 22). Já os cursos FIC+ correspondem a 36,4% (8 de 22).

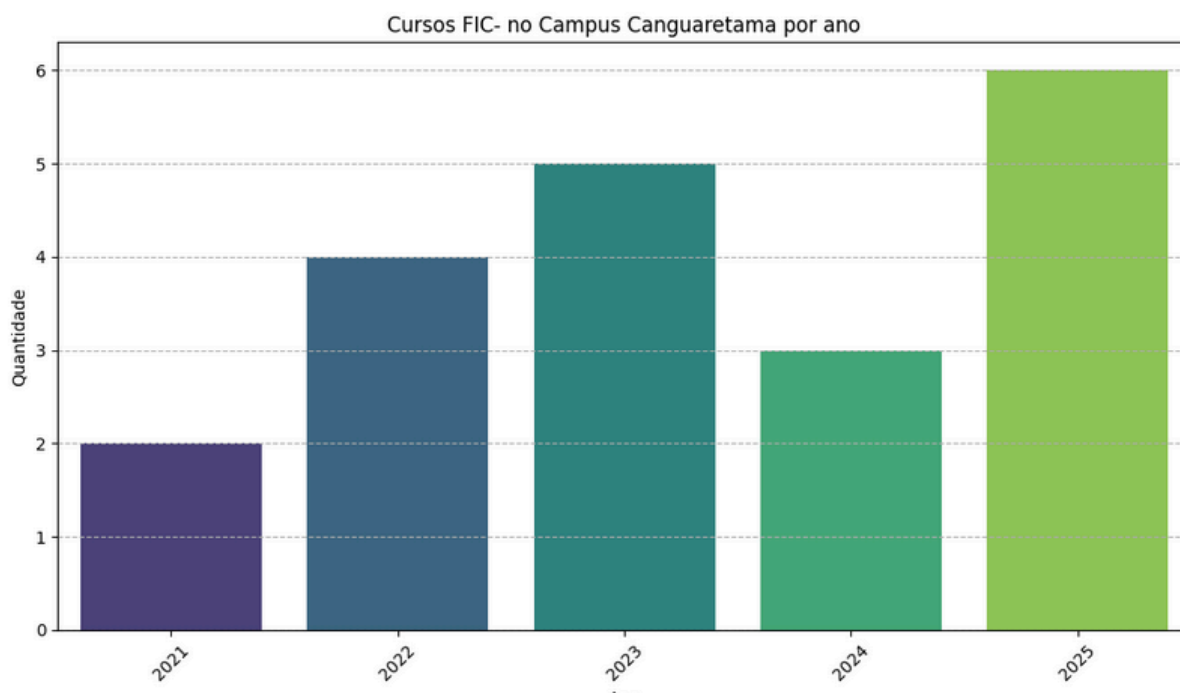
Em termos interpretativos, esse resultado reforça que, no período analisado, o Campus Canguaretama apresenta um perfil de oferta com predominância de cursos de menor carga horária, o que é compatível com a natureza da Formação Inicial e Continuada em muitos contextos. Ainda assim, a

presença de mais de um terço da oferta na categoria FIC+ indica que o campus também mantém participação relevante em cursos com carga horária mais elevada, geralmente associados a programas específicos ou formações mais estruturadas.

Por fim, é importante destacar que essa visualização considera cursos únicos (pela “Descrição do Curso”), ou seja, representa a diversidade de cursos ofertados, e não a quantidade de turmas abertas ou de matrículas realizadas.

A Figura 11 demonstra, por ano, a quantidade de cursos ofertados no Campus Canguaretama com carga horária inferior a 160 horas.

Figura 11. Cursos com carga horária inferior a 160 hrs no Campus Canguaretama por ano

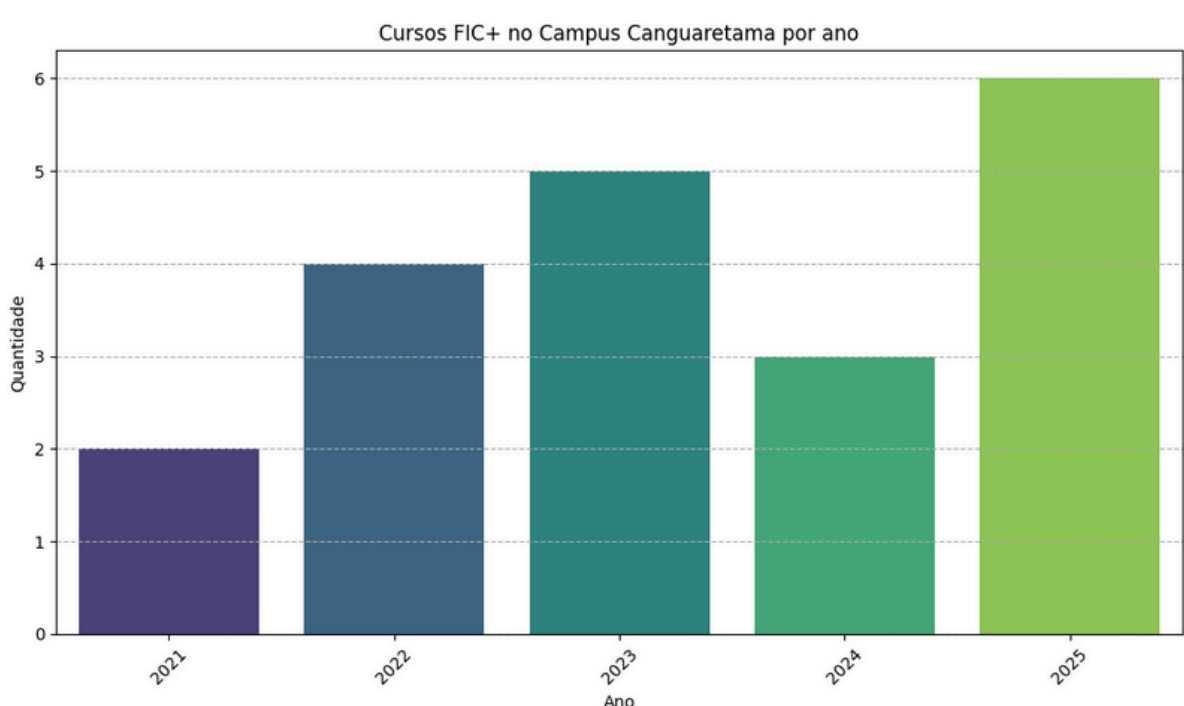


Fonte: Autoria própria(2026)

No Campus Canguaretama, os cursos FIC- apresentam um comportamento de expansão ao longo do período, com oscilações. Entre 2021 e 2023, há crescimento (de 2 para 5 cursos), seguido de redução em 2024 (3) e novo avanço em 2025, quando ocorre o maior registro (6). Em conjunto, o recorte indica aumento da diversidade de cursos de menor carga horária no campus, apesar de uma queda pontual no meio da série.

A Figura 12 apresenta a distribuição anual dos cursos ofertados no Campus Canguaretama com carga horária igual ou superior a 160 horas.

Figura 12. Cursos com carga horária igual ou superior a 160 hrs no Campus Canguaretama por ano



Fonte: Autoria própria(2026)

Nos cursos FIC+, observa-se um crescimento entre 2021 e 2023, passando de 2 para 5 cursos. Em 2024, há uma redução para 3 cursos, indicando uma retração temporária na oferta. Já em 2025, verifica-se uma retomada do crescimento, alcançando 6 cursos, o maior quantitativo do período analisado. Esse comportamento evidencia que a oferta de cursos com maior carga horária não se manteve estável após 2023, apresentando variações ao longo dos anos.

4. CONCLUSÃO

O objetivo geral do trabalho foi atingido, uma vez que foi realizada uma análise descritiva e exploratória dos cursos FIC ofertados pelo IFRN com base em dados disponibilizados pelo SUAP.

A partir da consolidação e preparação dos dados, foi possível identificar padrões relevantes do período de 2020 a 2026, como a predominância histórica de cursos FIC- e o crescimento mais expressivo de cursos FIC+ nos anos recentes.

Também foi possível aprofundar a análise por recortes institucionais, incluindo campus e turno, além de um panorama local do Campus Canguaretama, contribuindo para uma leitura mais objetiva do comportamento das ofertas e apoiando a transparência e a interpretação dos dados.

As tecnologias utilizadas foram essenciais para viabilizar o estudo com qualidade e reprodutibilidade. Os dados foram obtidos do SUAP e organizados no Google Colab, onde a linguagem Python permitiu automatizar a preparação e a análise.

Bibliotecas como Pandas e NumPy deram suporte à consolidação, limpeza e transformação do dataset, enquanto Matplotlib e Seaborn facilitaram a construção de gráficos para interpretação dos resultados.

Como trabalhos futuros, recomenda-se evoluir a solução para reduzir dependências de extrações manuais e aumentar a atualização dos dados. Uma melhoria importante é planejar uma integração direta com o SUAP (via rotinas automatizadas de coleta/atualização), garantindo periodicidade e consistência no carregamento.

Também é recomendável ampliar os indicadores da análise, incorporando métricas como evolução de matrículas, situação no curso, taxas de conclusão (quando os campos estiverem adequadamente preenchidos) e comparações mais detalhadas por campus/turno. Por fim, pode-se considerar integrações com serviços do ecossistema Google (por exemplo, publicação e atualização em planilhas, compartilhamento de relatórios e visualizações).

REFERÊNCIAS

ANACONDA. Libere inovação e valor de IA. Disponível em: <https://www.anaconda.com/>. Acesso em: 15 nov. 2025.

BELL, Peter; BEER, Brent. Introdução ao GitHub: Um guia que não é técnico. Novatec Editora, 2015.

CHIZZOLINI, Jônatas Domingues de Almeida. Estudo comparativo de bibliotecas python para visualização de dados: matplotlib e seaborn. 2023.

DEODATO, Felipe Augusto de Oliveira; BARIONI, Gabriel Paz da Silva. Indexação e recuperação de recursos educacionais abertos em repositórios suportados por licenças creative commons e versionamento de código. 2019. Faculdade de Computação e Informática, Universidade Presbiteriana Mackenzie, São Paulo, Brasil.

DOCUMENTAÇÃO do pandas. Disponível em: <https://pandas.pydata.org/docs/#module-pandas>. Acesso em: 15 nov. 2025.

HUNTER, J. D. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9(3), 90–95, 2007.

IFRN. Diretrizes orientadoras para os cursos de qualificação profissional (FIC). Pró-Reitoria de Ensino; Pró-Reitoria de Extensão. Natal: IFRN, 2025.

LOPES, Gesiel & ALMEIDA, Alessandro & DELBEM, Alexandre & TOLEDO, Claudio. Introdução à Análise Exploratória de Dados com Python. 2019.

LOPES, Gesiel Rios et al. Introdução à análise exploratória de dados com python. Minicursos ERCAS ENUCMPI, v. 2019, p. 160-176, 2019.

REIS, Edna Afonso; REIS, Ilka Afonso. Análise descritiva de dados. Relatório Técnico do Departamento de Estatística da UFMG, v. 1, 2002.

SILVESTRE, António. Análise de dados e estatística descritiva. Escolar Editora, 2007.

ANEXOS

Link para acesso ao notebook do Google Colab utilizado na
preparação e análise dos dados:

<https://colab.research.google.com/drive/1ZzFNrW8t53SpLNMOacG-6vmsOV9MSNL>