

**UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE
MESQUITA FILHO"**

**FACULDADE DE CIÊNCIAS - CAMPUS BAURU
DEPARTAMENTO DE COMPUTAÇÃO
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

LEONARDO FERREIRA ROCCATO

SISVIA: UMA FERRAMENTA DE APOIO À GESTÃO

**BAURU
Novembro/2025**

LEONARDO FERREIRA ROCCATO

SISVIA: UMA FERRAMENTA DE APOIO À GESTÃO

Trabalho de Conclusão de Curso do curso de Bacharelado em Ciência da Computação da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Ciências, Campus Bauru.

Orientadora: Profa. Dra. Simone das Graças Domingues Prado

BAURU
Novembro/2025

R671s Roccato, Leonardo
SISVIA: Uma ferramenta de apoio à gestão / Leonardo Roccato. --
Bauru, 2025
53 p. : il.

Trabalho de conclusão de curso (Bacharelado - Ciência da
Computação) - Universidade Estadual Paulista (UNESP), Faculdade
de Ciências, Bauru

Orientadora: Simone das Graças Domingues Prado

1. Visualização de Dados. 2. Indicadores Acadêmicos. 3. Gestão
Universitária. 4. Previsão de Séries Temporais. I. Título.

Leonardo Ferreira Roccato

SISVIA: UMA FERRAMENTA DE APOIO À GESTÃO

Trabalho de Conclusão de Curso do
curso de Bacharelado em Ciência da
Computação da Universidade
Estadual Paulista "Júlio de Mesquita
Filho", Faculdade de Ciências,
Campus Bauru.

Banca Examinadora

Simone das Graças Domingues Prado

Orientadora

Universidade Estadual Paulista "Júlio de Mesquita Filho"

Faculdade de Ciências

Departamento de Computação

Andréa Carla Gonçalves Vianna

Universidade Estadual Paulista "Júlio de Mesquita Filho"

Faculdade de Ciências

Departamento de Computação

Kleber Rocha de Oliveira

Universidade Estadual Paulista "Júlio de Mesquita Filho"

Faculdade de Ciências

Departamento de Computação

Bauru, 13 de Novembro de 2025.

Dedico este trabalho aos meus pais que me incentivaram a cursar a faculdade primeiramente e me apoiaram durante o período, e também aos professores e grandes amigos que fiz durante essa trajetória.

AGRADECIMENTOS

Agradeço especialmente a Profa. Dra. Simone das G Domingues Prado pela idealização do projeto de pesquisa e todo o suporte promovido durante sua execução e a Profa. Dra. Andréa Carla Gonçalves Vianna pelo papel decisivo para que eu possa ter chego até aqui, serei eternamente grato a professora que me ensinou a programar e que me ajudou e incentivou a concluir a graduação quando já não parecia mais possível

RESUMO

Este trabalho apresenta o desenvolvimento de um sistema de visualização de indicadores acadêmicos voltado ao apoio à gestão universitária no campus da UNESP Bauru. O sistema consolida e comunica métricas de ingressantes e formandos por curso, ano e tipo de convocação em um painel interativo, promovendo análises dinâmicas e decisões baseadas em dados. As informações foram coletadas a partir do Anuário Estatístico da UNESP e das estatísticas da VUNESP, abrangendo séries históricas de matrículas e formaturas. Metodologicamente, trata-se de um estudo aplicado e tecnológico, desenvolvido em Python com uso das bibliotecas pandas, NumPy, Altair e Streamlit. O fluxo implementado contempla tratamento e integração de planilhas institucionais, cálculo de indicadores de ocupação, distribuição por gênero, participação por modalidade de ingresso e cruzamento entre ingressantes e formandos. O sistema também incorpora um módulo de previsão temporal, utilizando os métodos Naive, Média Móvel, Holt e Prophet para estimar tendências futuras de ingresso e conclusão. Os resultados evidenciam que a ferramenta reduz o esforço manual, amplia a precisão analítica e facilita o acompanhamento de políticas acadêmicas, configurando-se como um instrumento eficaz, replicável e escalável para fortalecer a cultura de gestão baseada em evidências na universidade.

Palavras-chave: visualização de dados; indicadores acadêmicos; gestão universitária; previsão de séries temporais; Python; Streamlit; UNESP.

ABSTRACT

This study presents the development of an academic indicators visualization system designed to support university management at the UNESP Bauru campus. The system consolidates and communicates metrics of freshmen and graduates by course, year, and admission type through an interactive dashboard, enabling dynamic analyses and data-driven decision-making. The information was collected from the UNESP Statistical Yearbook and VUNESP's official statistics, encompassing historical data on enrollments and graduations. Methodologically, this is an applied and technological study developed in Python using the pandas, NumPy, Altair, and Streamlit libraries. The implemented workflow includes the treatment and integration of institutional spreadsheets, the calculation of indicators such as vacancy occupancy, gender distribution, admission modality participation, and the correlation between freshmen and graduates. The system also incorporates a time series forecasting module employing the Naive, Moving Average, Holt, and Prophet methods to estimate future trends in enrollments and completions. The results demonstrate that the tool reduces manual effort, enhances analytical accuracy, and facilitates monitoring of academic policies, proving to be an effective, replicable, and scalable instrument that strengthens the culture of evidence-based management within the university.

Keywords: data visualization; academic indicators; university management; time series forecasting.

LISTA DE FIGURAS

Figura 1 – Anuário Estatístico UNESP	17
Figura 2 – VUNESP Estatística de Vestibular	18
Figura 3 – Aplicativo: Filtros Interativos	24
Figura 4 – Aplicativo: Abas	24
Figura 5 – Aplicativo: Visão Geral Cabeçalho	25
Figura 6 – Aplicativo: Matrículas por Ano	25
Figura 7 – Aplicativo: Aba "Por Gênero"	25
Figura 8 – Aplicativo: Drilldown por Curso	26
Figura 9 – Aplicativo: Formandos por Ano	27
Figura 10 – Aplicativo: Formandos por Gênero	27
Figura 11 – Aplicativo: Série Histórica por Curso	27
Figura 12 – Aplicativo: Cruzado (Ingressantes x Formandos)	28
Figura 13 – Aplicativo: Previsões	29
Figura 14 – Aplicativo: Programa de Cotas	30
Figura 15 – Aplicativo: Série Histórica Programa de Cotas por Curso	31
Figura 16 – Aplicativo: Exportação de Dados	31
Figura 17 – Aplicativo: Programa de Cotas Caso de Estudo	32

LISTA DE ABREVIATURAS E SIGLAS

API	Application Programming Interface
CSV	Comma-Separated Values
IDE	Integrated Development Environment
KPI	Key Performance Indicator
MAPE	Mean Absolute Percentage Error
PPI	Pretos, Pardos e Indígenas
PROPEG	Pró-Reitoria de Planejamento Estratégico e Gestão
RMSE	Root Mean Square Error
SRVEBP	Sistema de Reserva de Vagas para Educação Básica Pública
SU	Sistema Universal
UNESP	Universidade Estadual Paulista “Júlio de Mesquita Filho”
VUNESP	Fundação para o Vestibular da Universidade Estadual Paulista

Sumário

1. INTRODUÇÃO	13
1.1 Problemática	14
1.2 Justificativa	15
1.3 Objetivos	16
1.3.1 Objetivo geral	16
1.3.2 Objetivos específicos	17
2. FUNDAMENTAÇÃO TEÓRICA	18
2.1. Gestão Universitária	18
2.2. Visualização de Dados e Sistemas Interativos	18
2.3. Modelos de Previsão Aplicados a Séries Temporais	19
3. MATERIAIS E MÉTODOS	21
3.1 Metodologia	21
3.1.1 Tipo e Abordagem de Pesquisa	21
3.1.2 Fontes de Dados	21
3.1.3 Etapas do Desenvolvimento	22
3.1.4 Critérios de Desenvolvimento	23
3.2 Tecnologias Utilizadas	23
3.3. Ambiente de Desenvolvimento	23
3.4. Bibliotecas Utilizadas	24
3.5. Hardware Utilizado	26
3.6. Implementação via Ambiente Streamlit Cloud	26
4. DOCUMENTAÇÃO DE SOFTWARE	28
4.1. Fontes de Dados	28
4.1.1. Anuário Estatístico UNESP	28
4.1.2. VUNESP	29
4.2. Extração dos Dados	30
4.3. Análise de Requisitos	31
4.3.1. Requisitos Funcionais	31

4.3.2. Requisitos Não-Funcionais	32
4.4. Base de Dados-----	32
4.5. Estrutura e Arquitetura do Código -----	33
4.6. Carregamento e Tratamento dos Dados-----	33
4.7. Geração dos Indicadores e Métricas-----	34
4.8. Interface do Dashboard (Streamlit)-----	35
4.9. Menu lateral e filtros interativos -----	35
4.10. Abas Principais do Sistema -----	36
4.10.1. Aba “Visão Geral” -	36
4.10.2. Aba “Por Gênero” -	37
4.10.3. Aba “Por Curso” -	38
4.10.4. Aba “Formandos” -	38
4.10.5. Aba “Cruzado: Ingressantes x Formandos” -----	40
4.10.6. Aba “Previsões” -	41
4.10.7. Aba “Programa de Cotas” -	42
4.10.8. Exportação e Atualização dos Dados-----	43
4.11. Módulo de Previsões e Análises Temporais-----	44
4.12. Avaliação de desempenho e usabilidade do sistema -----	46
4.13. Estudo de Caso: Evolução do Programa de Cotas na UNESP -----	46
4.14. Discussão dos resultados obtidos-----	48
5. CONCLUSÃO -----	50
REFERÊNCIAS-----	52

1. INTRODUÇÃO

Nas últimas décadas, a transformação digital tem impulsionado mudanças significativas no modo como as instituições de ensino superior gerenciam e utilizam seus dados acadêmicos. A disponibilidade crescente de informações institucionais — provenientes de sistemas de matrícula, anuários estatísticos e relatórios de desempenho — oferece um potencial expressivo para aprimorar os processos de planejamento e tomada de decisão. No entanto, a ausência de ferramentas integradas e acessíveis de análise e visualização ainda representa um desafio para gestores e pesquisadores das universidades públicas brasileiras.

No contexto da UNESP, a gestão de indicadores de ingresso, matrícula e conclusão dos cursos apresenta alto volume de dados provenientes de diferentes fontes, como o Anuário Estatístico da UNESP (UNESP, 2025) e os relatórios da Fundação VUNESP (VUNESP, 2025). A análise manual dessas informações demanda tempo, conhecimentos técnicos específicos e pode resultar em inconsistências entre relatórios. Tais dificuldades evidenciam a necessidade de uma solução tecnológica que centralize, padronize e apresente essas informações de maneira clara, interativa e confiável.

O uso de técnicas de análise de dados e visualização interativa constitui uma estratégia eficaz para transformar dados brutos em informações compreensíveis e úteis à gestão. Nesse sentido, ferramentas desenvolvidas em Python (VAN ROSSUM; DRAKE, 2009), com suporte de bibliotecas como pandas (MCKINNEY, 2010), NumPy (OLIPHANT, 2006), Altair (VANDERPLAS, 2018) e Streamlit (STREAMLIT INC., 2023), permitem construir sistemas capazes de manipular grandes volumes de dados, gerar gráficos dinâmicos e apresentar indicadores em tempo real. Quando associadas a métodos estatísticos de previsão, como Naive, Média Móvel, Holt e Prophet, essas tecnologias ampliam o potencial analítico, possibilitando estimativas de tendências futuras e o acompanhamento evolutivo de cursos e modalidades de ingresso.

Diante desse cenário, este trabalho tem como propósito desenvolver um sistema de visualização de indicadores acadêmicos voltado ao apoio à gestão universitária no campus da UNESP Bauru. O sistema propõe a consolidação de dados históricos de ingressantes e formandos, a geração de métricas analíticas e o cruzamento entre séries temporais, permitindo que gestores e pesquisadores compreendam padrões, comparem desempenhos e realizem projeções de evolução acadêmica.

A relevância desta pesquisa está na contribuição para a modernização da gestão universitária e na promoção de uma cultura institucional orientada por dados. Ao automatizar processos de coleta, análise e visualização, a ferramenta proposta busca reduzir o tempo gasto em tarefas operacionais, aumentar a precisão das informações e favorecer a transparência na divulgação de resultados acadêmicos.

1.1 Problemática

A administração universitária contemporânea demanda decisões cada vez mais fundamentadas em dados quantitativos e evidências objetivas. Entretanto, nas universidades públicas brasileiras, grande parte das informações acadêmicas permanece dispersa entre relatórios, planilhas e sistemas administrativos heterogêneos, dificultando a análise integrada dos indicadores institucionais. Essa fragmentação compromete a capacidade de gestores, coordenadores e pesquisadores de acompanhar a evolução de cursos, identificar gargalos e planejar estratégias de melhoria com base em dados consolidados.

No caso da UNESP, o Anuário Estatístico e as estatísticas divulgadas pela VUNESP reúnem um conjunto valioso de informações sobre ingressos, matrículas e formaturas. Contudo, o formato predominantemente tabular e estático desses relatórios exige tratamento manual para cruzamento, visualização e interpretação das informações. Tal prática é suscetível a erros, retrabalho e inconsistências entre análises, especialmente quando os dados abrangem diversos cursos e períodos históricos.

A ausência de uma ferramenta integrada para visualização de indicadores acadêmicos limita a capacidade da instituição de extrair insights relevantes e

acompanhar o desempenho de suas políticas de acesso e permanência. Além disso, a inexistência de mecanismos automatizados de previsão impede o planejamento proativo de metas e ações de melhoria. Métodos estatísticos e computacionais como Naive, Média Móvel, Holt e Prophet poderiam antecipar tendências de ingresso e conclusão, fornecendo subsídios valiosos para o dimensionamento de vagas, análise de evasão e avaliação de políticas de inclusão.

Dessa forma, identifica-se como problema central deste trabalho a carência de um sistema que integre, analise e apresente de forma dinâmica os indicadores acadêmicos da UNESP, permitindo também projeções futuras baseadas em séries temporais. A inexistência dessa integração dificulta o uso estratégico das informações disponíveis, restringindo a eficiência das decisões de gestão e a transparência institucional.

1.2 Justificativa

A adoção de práticas de gestão baseadas em evidências tem se tornado um elemento essencial para a eficiência das instituições de ensino superior. No entanto, apesar do grande volume de dados disponíveis nas universidades públicas, o uso efetivo dessas informações ainda é limitado por barreiras técnicas e estruturais. Nesse contexto, a criação de um sistema de visualização de indicadores acadêmicos representa uma contribuição relevante tanto do ponto de vista científico quanto institucional, ao propor um modelo de integração, análise e comunicação de dados voltado ao apoio à gestão.

No caso da UNESP, os dados acadêmicos estão dispersos em diferentes plataformas, como o Anuário Estatístico e as estatísticas da VUNESP, exigindo esforços consideráveis para consolidação e interpretação. A ausência de uma ferramenta unificada dificulta a comparação histórica e o acompanhamento de indicadores fundamentais, tais como taxas de matrícula, conclusão e distribuição por modalidade de ingresso. Assim, há uma lacuna significativa entre a disponibilidade dos dados e sua utilização prática nos processos decisórios da universidade.

A relevância deste trabalho também se sustenta na necessidade de democratizar o acesso à informação institucional, tornando-a compreensível e acessível a gestores, docentes e pesquisadores, mesmo sem conhecimento técnico

em análise de dados. A utilização de tecnologias de código aberto como Python, Pandas, Altair e Streamlit reforça o caráter sustentável e replicável da solução, possibilitando sua expansão para outros campi e contextos educacionais.

Outro ponto de destaque é a incorporação de modelos preditivos no sistema, com base em métodos como Naive, Média Móvel, Holt (HOLT, 1957) e Prophet (TAYLOR; LETHAM, 2018), que permitem gerar previsões sobre o comportamento futuro de indicadores como ingressos e formaturas. Essa abordagem fornece uma visão prospectiva do cenário acadêmico, apoiando o planejamento de políticas institucionais e a alocação de recursos de forma mais estratégica.

Sob o ponto de vista científico, o projeto contribui para a área de Ciência de Dados aplicada à gestão universitária, ao integrar técnicas de análise estatística, visualização interativa e aprendizado de máquina em um mesmo ambiente. Do ponto de vista prático, promove transparência, agilidade e confiabilidade no uso dos dados acadêmicos, fortalecendo a governança informacional da universidade.

Portanto, a justificativa deste trabalho reside na necessidade de transformar dados dispersos e pouco explorados em conhecimento estratégico, consolidando um instrumento tecnológico que favoreça a gestão baseada em evidências, a transparência institucional e a eficiência na tomada de decisões dentro do ambiente universitário.

1.3 Objetivos

O presente trabalho tem como finalidade propor o desenvolvimento de um sistema computacional voltado à análise e visualização de indicadores acadêmicos da UNESP, campus de Bauru. Essa iniciativa surge da necessidade de integrar informações dispersas em diferentes relatórios institucionais, transformando-as em conhecimento acessível e útil para a gestão universitária. A seção a seguir apresenta os objetivos que orientam esta pesquisa, sendo o primeiro de caráter geral, que define o propósito central do estudo e os demais de natureza específica, que detalham as etapas e metas necessárias para o alcance do objetivo principal.

1.3.1 Objetivo geral

Desenvolver um sistema de visualização de indicadores acadêmicos que

integre, analise e apresente dados de ingressantes e formandos da UNESP Bauru de forma interativa e dinâmica, possibilitando a extração de informações relevantes e a geração de previsões que apoiem a gestão universitária baseada em evidências.

1.3.2 Objetivos específicos

Para atingir o objetivo geral, este trabalho propõe inicialmente consolidar os dados acadêmicos provenientes do *Anuário Estatístico da UNESP* e das estatísticas disponibilizadas pela *VUNESP*, organizando-os em um formato padronizado e compatível com o processamento computacional. Em seguida, serão implementadas rotinas em Python voltadas ao tratamento, integração e padronização das informações, de modo a garantir a consistência e a reprodutibilidade dos resultados.

Com base nesses dados estruturados, pretende-se calcular e analisar indicadores como taxa de ocupação de vagas, distribuição por gênero, por curso e formandos considerando diferentes cursos e períodos históricos. Tais indicadores serão apresentados em painéis interativos desenvolvidos na biblioteca *Streamlit*, utilizando *Altair* para a criação de gráficos dinâmicos e intuitivos.

O sistema também contemplará um módulo específico para o cruzamento entre ingressantes e formandos, permitindo observar tendências temporais e correlações entre o número de estudantes que ingressam e concluem cada curso. Além disso, serão aplicados modelos estatísticos de previsão temporal, como Naive, Média Móvel, Holt e Prophet, com o intuito de estimar a evolução futura dos indicadores acadêmicos e antecipar cenários de ingresso e conclusão.

Por fim, o projeto buscará validar o sistema por meio de testes exploratórios, avaliando sua usabilidade e clareza visual, bem como documentar toda a arquitetura e o fluxo de desenvolvimento, de modo a possibilitar sua replicação e expansão para outros campi da UNESP ou instituições que enfrentem desafios semelhantes de gestão e análise de dados acadêmicos.

2. FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica tem como propósito apresentar os conceitos e estudos que embasam o desenvolvimento deste trabalho, estabelecendo a ligação entre as bases científicas da pesquisa e a aplicação prática proposta. São abordados, a seguir, temas relacionados à gestão universitária orientada por dados, à visualização de informações e às técnicas de previsão aplicadas a séries temporais, além das tecnologias utilizadas no desenvolvimento do sistema.

2.1. Gestão Universitária

A gestão universitária moderna tem se apoiado cada vez mais em indicadores quantitativos para orientar decisões relacionadas a planejamento, avaliação institucional e políticas de permanência estudantil. Segundo Drucker (1999), “o que não é medido não pode ser gerenciado”, e esse princípio é particularmente aplicável ao contexto acadêmico, em que decisões administrativas impactam diretamente a qualidade do ensino e a eficiência dos recursos públicos.

De acordo com Fialho e Silva (2020), a adoção de modelos de gestão baseados em evidências permite identificar gargalos institucionais e acompanhar metas estratégicas de forma contínua. No caso das universidades públicas, a integração de dados de ingresso, evasão e conclusão é essencial para avaliar a efetividade das políticas de acesso e inclusão.

No contexto da UNESP, a utilização de relatórios e anuários estatísticos fornece informações valiosas sobre o desempenho acadêmico, mas ainda carece de ferramentas que permitam a exploração interativa desses dados. Assim, sistemas automatizados de análise podem contribuir significativamente para a transparência e eficiência administrativa, aproximando a universidade de modelos de gestão baseados em dados (*Data-Driven Management*).

2.2. Visualização de Dados e Sistemas Interativos

A visualização de dados é uma área interdisciplinar que combina estatística, design e ciência da computação para transformar dados complexos em

representações visuais compreensíveis. Conforme Tufte (2001), o objetivo da visualização é “comunicar informações de forma clara e eficiente”, permitindo ao observador identificar padrões, relações e anomalias.

Ferramentas de visualização interativa têm ganhado destaque por facilitar o acesso à informação e ampliar a capacidade analítica dos usuários. Segundo Few (2012), painéis (*dashboards*) bem projetados sintetizam dados relevantes e reduzem o tempo necessário para a interpretação de métricas complexas.

No ambiente acadêmico, sistemas desse tipo são especialmente úteis para acompanhar indicadores institucionais, uma vez que permitem filtrar informações por curso, ano ou categoria de ingresso, promovendo análises comparativas e históricas. O uso de tecnologias de código aberto, como Python, Streamlit e Altair, torna possível a criação de plataformas visuais flexíveis, de baixo custo e alta reprodutibilidade.

2.3. Modelos de Previsão Aplicados a Séries Temporais

A análise de séries temporais consiste em estudar o comportamento de uma variável ao longo do tempo, identificando tendências, ciclos e possíveis padrões de sazonalidade. Essa abordagem permite compreender o passado para estimar valores futuros com base nas relações temporais observadas nos dados (HYNDMAN; ATHANASOPOULOS, 2018).

No contexto deste trabalho, foram empregados métodos estatísticos e heurísticos de previsão capazes de gerar estimativas de curto prazo de forma simples e interpretável. Entre eles, destacam-se os modelos Naive, Média Móvel, Holt e Prophet.

O método Naive é considerado o modelo mais básico de previsão, assumindo que o valor futuro será igual ao último valor observado. Apesar de sua simplicidade, ele serve como uma referência inicial (*baseline*) para avaliar a qualidade de métodos mais avançados. Já a Média Móvel de três períodos utiliza a média aritmética dos últimos valores históricos para suavizar variações aleatórias e projetar tendências de estabilidade.

O modelo Holt, proposto por Charles Holt (1957), introduz o conceito de suavização exponencial dupla, incorporando um componente de tendência linear. Essa técnica é adequada para séries que apresentam crescimento ou declínio gradual

ao longo do tempo, oferecendo previsões mais realistas do que os modelos puramente estáticos.

Por fim, o modelo Prophet, desenvolvido por Taylor e Letham (2018) no Facebook (atual Meta), representa uma abordagem moderna e automatizada para previsão de séries temporais. Ele utiliza decomposição aditiva para modelar tendência, sazonalidade e efeitos residuais, ajustando-se facilmente mesmo a séries curtas ou irregulares. Além disso, fornece intervalos de confiança que expressam a incerteza associada às previsões, aumentando a interpretabilidade dos resultados.

A aplicação conjunta desses métodos permite avaliar diferentes perspectivas da evolução dos indicadores de ingressantes e formandos da UNESP, fornecendo subsídios quantitativos para o planejamento de políticas de acesso, permanência e gestão acadêmica baseada em dados.

3. MATERIAIS E MÉTODOS

3.1 Metodologia

A metodologia deste trabalho caracteriza-se como de natureza aplicada, com abordagem quantitativa e método experimental de desenvolvimento tecnológico, voltado à criação de uma ferramenta computacional que transforma dados acadêmicos em informações estratégicas para a gestão universitária. O estudo não se limita à análise de dados existentes, mas propõe a implementação de um sistema interativo que integra, visualiza e prevê indicadores institucionais, configurando um produto técnico de relevância prática e científica.

3.1.1 Tipo e Abordagem de Pesquisa

A pesquisa é classificada como aplicada, pois busca gerar soluções diretas para um problema real de gestão de dados na UNESP. Adota uma abordagem quantitativa, uma vez que se fundamenta na coleta, tratamento e análise de dados numéricos referentes a ingressantes e formandos, buscando identificar padrões e tendências. Em termos de procedimentos técnicos, trata-se de uma pesquisa de desenvolvimento, pois envolve a concepção e a implementação de um sistema computacional funcional, que é posteriormente avaliado quanto à sua eficiência e aplicabilidade.

3.1.2 Fontes de Dados

As informações utilizadas na pesquisa foram coletadas a partir de fontes oficiais e públicas da própria universidade, em especial o *Anuário Estatístico da UNESP* (UNESP, 2025), divulgado no Portal Institucional da Universidade Estadual Paulista, documento que compila dados consolidados a respeito dos cursos de graduação, como quantidade de vagas disponíveis, matrículas realizadas e total de alunos formados., e as estatísticas de vestibular da Fundação VUNESP (VUNESP, 2025), na seção institucional dedicada à divulgação de resultados, indicadores e informações numéricas relacionadas ao processo seletivo da universidade. Os dados abrangem séries históricas de cursos de graduação, contendo informações sobre número de

vagas, matrículas efetivadas, gênero dos ingressantes e número de formandos por ano.

3.1.3 Etapas do Desenvolvimento

O processo metodológico adotado pode ser dividido em cinco etapas principais: coleta e organização dos dados, tratamento e padronização, análise e cálculo dos indicadores, desenvolvimento do sistema de visualização e implementação dos modelos preditivos.

Na primeira etapa, realizou-se a importação das planilhas do Anuário e da VUNESP, consolidando as informações em uma base única, garantindo a compatibilidade entre os conjuntos de dados. Em seguida, na etapa de tratamento, os dados passaram por um processo de limpeza, com a correção de inconsistências, normalização de nomenclaturas e preenchimento de valores faltantes, utilizando as bibliotecas pandas e numpy.

Na terceira etapa, foram criadas métricas analíticas que permitem mensurar a evolução dos cursos, incluindo a taxa de ocupação de vagas, a distribuição por sexo e a participação por tipo de ingresso. Esses indicadores foram estruturados em funções específicas no código-fonte, garantindo modularidade e reuso.

A quarta etapa consistiu no desenvolvimento da interface interativa, utilizando a biblioteca Streamlit, que permite criar painéis web dinâmicos em Python. Foram implementados filtros por curso, ano e tipo de convocação, além de gráficos desenvolvidos com a biblioteca Altair, que permitem visualizar séries temporais e comparações entre grupos de dados.

Por fim, na quinta etapa, o sistema foi ampliado com a inclusão de um módulo de previsão temporal, no qual foram aplicados os modelos Naive, Média Móvel, Holt e Prophet, para estimar o comportamento futuro de indicadores como número de ingressantes e formandos. Esses modelos foram ajustados e avaliados com base em métricas de erro e validação cruzada, assegurando a coerência das projeções geradas.

3.1.4 Critérios de Desenvolvimento

A avaliação do sistema ocorreu de forma exploratória, considerando critérios de

funcionalidade, usabilidade e clareza das visualizações. Foram analisados o tempo de resposta da aplicação, a coerência entre os indicadores calculados e os dados originais, além da capacidade de o painel gerar insights interpretáveis por diferentes perfis de usuários. Também foram realizadas simulações com séries históricas para verificar a robustez dos modelos preditivos e sua adequação à variação temporal dos indicadores.

3.2 Tecnologias Utilizadas

A escolha da linguagem Python deve-se à sua versatilidade, ampla comunidade científica e integração com bibliotecas de análise de dados. O pandas e o numpy foram empregados para manipulação e limpeza das bases, enquanto Altair e matplotlib foram utilizados para geração de visualizações gráficas. A biblioteca Streamlit foi adotada como plataforma de interface interativa, permitindo a execução do sistema diretamente em ambiente web.

A estrutura modular do código-fonte favorece a escalabilidade e a reprodutibilidade dos resultados, permitindo a atualização dos dados com facilidade e a extensão do sistema para novas métricas ou bases institucionais. Essa arquitetura está em consonância com o conceito de ciência aberta e reprodutível, defendido por Peng (2011), que preconiza a transparência dos métodos analíticos como parte integrante da confiabilidade científica.

Dessa forma, o conjunto de tecnologias e fundamentos teóricos apresentados nesta seção constitui a base conceitual para o desenvolvimento e a validação do sistema de visualização de indicadores acadêmicos proposto neste trabalho.

3.3. Ambiente de Desenvolvimento

O desenvolvimento do sistema foi realizado no ambiente PyCharm, da JetBrains, uma IDE amplamente utilizada para projetos em Python.

O PyCharm oferece recursos como gerenciamento de pacotes integrados, ambiente virtual automatizado (venv), depuração passo a passo, controle de versão via Git e suporte a *frameworks* de dados, o que facilita a manutenção e a escalabilidade do código.

Além disso, sua integração com o Streamlit permitiu a execução e o teste do dashboard diretamente na IDE, agilizando o ciclo de desenvolvimento e reduzindo erros de configuração.

3.4. Bibliotecas Utilizadas

Com base no histórico do projeto, as principais bibliotecas empregadas no desenvolvimento do sistema foram:

A biblioteca pandas constitui o núcleo do processamento de dados do sistema. Desenvolvida para análise e manipulação eficiente de estruturas tabulares, ela permite importar planilhas em formato .xlsx, realizar filtragens, agregar informações e corrigir inconsistências. No projeto, o pandas foi utilizado para ler as bases do Anuário Estatístico e da VUNESP, tratar colunas numéricas e textuais e estruturar os dados em *DataFrames*, que servem de entrada para os módulos analíticos e de visualização. Sua sintaxe expressiva e integração com outras bibliotecas do ecossistema Python tornam-na essencial para a transformação de dados brutos em conjuntos prontos para análise.

A biblioteca NumPy (*Numerical Python*) foi empregada para cálculos numéricos e manipulação de vetores e matrizes, oferecendo operações vetorizadas de alta performance. Suas estruturas de dados otimizadas e funções matemáticas foram utilizadas para cálculos de taxas, médias móveis e projeções. Além disso, o NumPy atua como base para diversas outras bibliotecas empregadas no projeto — como *pandas* e *statsmodels* —, fornecendo suporte para operações matemáticas complexas e manipulação eficiente de arrays.

A biblioteca Altair foi responsável pela geração das visualizações interativas e estatísticas do sistema. Baseada no *framework Vega-Lite*, ela permite criar gráficos declarativos e altamente personalizáveis, mantendo clareza e estética visual. No projeto, o *Altair* foi utilizado para construir gráficos de linhas, barras e áreas que representam séries temporais, comparações entre cursos e distribuições percentuais. Sua integração com o *Streamlit* possibilitou que as visualizações fossem renderizadas diretamente na interface web, atualizando-se automaticamente de acordo com os filtros aplicados pelo usuário.

A biblioteca Streamlit foi utilizada como plataforma principal de interface e

execução do sistema. Projetada para transformar scripts Python em aplicações web interativas com rapidez, ela permitiu que o sistema funcionasse como um painel analítico acessível via navegador, sem necessidade de servidores complexos ou frameworks de front-end. O *Streamlit* gerencia o carregamento dos dados, os filtros laterais, as abas de navegação e a renderização dos gráficos e tabelas. Essa escolha garantiu simplicidade na implantação, usabilidade e escalabilidade, características essenciais para o contexto acadêmico e institucional da UNESP.

A biblioteca *statsmodels* foi empregada para implementação dos modelos estatísticos de previsão, especialmente o modelo de Holt, que utiliza suavização exponencial para capturar tendências lineares. Essa biblioteca fornece funções para análise de séries temporais, estimação de modelos e cálculo de métricas de desempenho. No projeto, ela foi fundamental para o módulo de previsões, permitindo que o sistema gerasse projeções de curto prazo com base nos dados históricos de ingressantes e formandos.

Desenvolvida pela Meta (Facebook), a biblioteca *Prophet* foi utilizada como um dos principais métodos de previsão temporal do sistema. Seu algoritmo combina decomposição aditiva de tendência, sazonalidade e ruído, oferecendo resultados precisos e interpretáveis mesmo para séries curtas ou irregulares. O *Prophet* permite também a geração automática de intervalos de confiança, facilitando a avaliação da incerteza das projeções. Dentro do painel, ele foi aplicado para estimar cenários futuros de ingresso e conclusão de cursos, ampliando o potencial analítico e preditivo da ferramenta.

A biblioteca *Matplotlib* foi utilizada de forma complementar para geração de visualizações simples e testes de coerência gráfica durante o desenvolvimento. Embora o *Altair* tenha sido a principal ferramenta de visualização, o *Matplotlib* foi essencial em etapas de verificação, criação de gráficos rápidos e inspeção exploratória dos resultados, garantindo consistência entre as representações gráficas e os dados originais.

A biblioteca *scikit-learn* foi empregada em experimentos comparativos de modelagem preditiva e validação dos resultados obtidos pelos métodos implementados. Ela oferece um conjunto abrangente de algoritmos de aprendizado supervisionado e não supervisionado, além de métricas estatísticas, incluindo o cálculo de RMSE e MAPE, utilizados para avaliar o desempenho dos modelos de

previsão. Embora não tenha sido o núcleo do sistema, o *scikit-learn* foi importante para garantir a confiabilidade das avaliações de erro e a coerência estatística das projeções geradas.

Algumas bibliotecas de suporte foram empregadas para tarefas complementares. A *openpyxl* foi utilizada para leitura e gravação de planilhas Excel, permitindo que os arquivos *.xlsx* fossem manipulados diretamente pelo Python. As bibliotecas *os* e *pathlib* deram suporte à manipulação de diretórios, facilitando o acesso e a organização dos arquivos dentro do ambiente de execução. Já *time* e *datetime* foram utilizadas para o controle de execução, marcação temporal das séries históricas e automação de rotinas dependentes de data. Embora menos complexas, essas bibliotecas desempenharam papel essencial na robustez e estabilidade do sistema.

3.5. Hardware Utilizado

O desenvolvimento e os testes do sistema foram realizados em um Samsung Galaxy Book3 360, equipado com processador Intel Core i5- 1335U (12ª geração), 16 GB de memória RAM e SSD de 512 GB, executando o sistema operacional Windows 11 Home.

Esse hardware proporcionou desempenho adequado para a execução de modelos estatísticos, renderização de gráficos interativos e testes simultâneos da aplicação web.

A escolha de um notebook híbrido, com tela sensível ao toque e arquitetura eficiente, também possibilitou a realização de testes de usabilidade diretamente no ambiente de execução, simulando o comportamento esperado de usuários em dispositivos variados.

3.6. Implementação via Ambiente Streamlit Cloud

Para disponibilizar o sistema de visualização de indicadores acadêmicos de forma pública e acessível, optou-se pela implantação na plataforma Streamlit Cloud, serviço oficial de hospedagem da biblioteca *Streamlit*. Essa solução permite publicar aplicações desenvolvidas em Python diretamente na nuvem, sem necessidade de

infraestrutura própria, servidores dedicados ou conhecimento avançado em configuração de ambientes web.

O processo de implantação ocorre de maneira automatizada: o desenvolvedor armazena o código-fonte do projeto em um repositório GitHub, e o Streamlit Cloud realiza a integração contínua, instalando as dependências listadas no arquivo *requirements.txt* e executando o script principal (*app.py*) para gerar a interface interativa. Dessa forma, o sistema pode ser acessado instantaneamente por qualquer usuário por meio de um link público, sem necessidade de download, instalação local ou ambiente configurado.

Além da praticidade, o Streamlit Cloud apresenta vantagens significativas para fins acadêmicos e institucionais. Trata-se de uma plataforma gratuita para projetos de pequeno e médio porte, com capacidade suficiente para hospedar aplicações de análise de dados e painéis interativos, como o desenvolvido neste trabalho. O serviço oferece ainda atualizações automáticas, permitindo que qualquer modificação feita no repositório GitHub seja refletida imediatamente na aplicação online, garantindo sincronização constante e facilidade de manutenção.

A escolha do Streamlit Cloud também está alinhada aos princípios da ciência aberta e reproduzível, uma vez que o código-fonte e o painel podem ser disponibilizados publicamente, assegurando transparência e acesso democrático aos resultados. Dessa forma, gestores, docentes e pesquisadores podem acessar o sistema a qualquer momento, a partir de qualquer dispositivo com conexão à internet, reforçando o objetivo central do projeto de tornar os indicadores acadêmicos da UNESP mais acessíveis, atualizados e interpretáveis.

Assim, o uso do Streamlit Cloud mostrou-se uma alternativa eficiente, sustentável e de baixo custo para a disponibilização da ferramenta, consolidando o caráter prático e replicável da solução proposta e permitindo sua expansão futura para outros campi ou contextos institucionais.

4. DOCUMENTAÇÃO DE SOFTWARE

A presente seção descreve o funcionamento, a estrutura e os resultados obtidos com o desenvolvimento do sistema de visualização de indicadores acadêmicos proposto neste trabalho. São abordadas a arquitetura geral, o tratamento dos dados, o funcionamento das abas do painel, os métodos preditivos empregados e a avaliação de desempenho e usabilidade do sistema.

4.1. Fontes de Dados

Os dados utilizados neste trabalho foram obtidos de fontes institucionais oficiais, que fornecem informações consolidadas e públicas sobre o desempenho acadêmico da UNESP. Essas fontes são fundamentais para garantir a confiabilidade e a legitimidade das análises realizadas, além de assegurarem a reprodutibilidade do estudo.

As principais fontes consultadas foram o Anuário Estatístico da UNESP e as Estatísticas de Vestibular da Fundação VUNESP, que em conjunto oferecem uma visão abrangente sobre os processos de ingresso, matrícula e formatura nos cursos de graduação. Ambas as bases são disponibilizadas periodicamente, em formato aberto e padronizado, permitindo seu uso em estudos acadêmicos e aplicações institucionais.

4.1.1. Anuário Estatístico UNESP

O *Anuário Estatístico da UNESP (Figura 1)* é uma publicação oficial elaborada pela PROPEG, contendo indicadores quantitativos e qualitativos sobre a universidade.

Esse documento apresenta informações detalhadas sobre corpo discente, docente e técnico-administrativo, além de dados relativos a ingressantes, matrículas ativas, concluintes, evasão e infraestrutura.

No contexto deste trabalho, o anuário foi utilizado como principal fonte de dados sobre ingressantes e formandos da UNESP Bauru, abrangendo um período de vários anos letivos. As planilhas, disponíveis em formato .xlsx e .csv, foram extraídas diretamente do portal institucional da UNESP, permitindo posterior integração com

outras fontes.

Sua utilização assegura a confiabilidade dos indicadores e reflete as estatísticas oficiais utilizadas pela própria universidade em relatórios de gestão e avaliação institucional.

Figura 1 - Anuário Estatístico UNESP



Fonte: Anuário Estatístico UNESP (<https://www2.unesp.br/porta/#/anuario>)

4.1.2. VUNESP

A Fundação VUNESP é a instituição responsável pela organização e execução dos processos seletivos da UNESP.

As estatísticas divulgadas pela VUNESP (Figura 2) contemplam informações sobre candidatos inscritos, aprovados, matriculados e convocados, bem como dados desagregados por modalidade de ingresso, gênero, curso e unidade universitária.

Esses relatórios públicos foram utilizados para complementar o conjunto de dados do anuário, especialmente no que se refere à análise do programa de cotas e modalidades de convocação (ampla concorrência, isentos e convênios).

A base da VUNESP apresenta um formato altamente estruturado, o que facilitou sua leitura e padronização via Python. Além disso, o cruzamento das informações com o anuário estatístico permitiu validar os dados e eliminar possíveis divergências, resultando em uma base consolidada para o painel de indicadores.

Figura 2 - VUNESP Estatística de Vestibular



Fonte: VUNESP Estatística de Vestibular
(<https://www.vunesp.com.br/institucional/estatisticavestibular>)

4.2. Extração dos Dados

O processo de extração dos dados foi realizado de forma automatizada, utilizando scripts desenvolvidos em Python, com o objetivo de otimizar a coleta e o tratamento das informações provenientes dos relatórios institucionais da UNESP e da VUNESP.

Os relatórios do anuário, geralmente disponibilizados em formato PDF, foram processados por meio da biblioteca Camelot, que permite a leitura e conversão de tabelas em estruturas manipuláveis de dados, como *DataFrames*.

Após a extração inicial, as tabelas foram convertidas em arquivos .xlsx e organizadas por curso e ano correspondente.

Em seguida, empregou-se a biblioteca pandas para tratamento e integração das planilhas, abrangendo operações de limpeza (remoção de valores nulos e duplicados), normalização de nomes de cursos e adequação de tipos de dados numéricos.

Esse processo automatizado garante reprodutibilidade e rastreabilidade, permitindo que futuras atualizações das bases sejam incorporadas ao sistema com o mínimo de esforço manual.

Além disso, o código de extração foi projetado de forma modular, podendo ser adaptado para novas fontes ou formatos, o que contribui para a escalabilidade do sistema e sua aplicação em outras unidades universitárias.

4.3. Análise de Requisitos

A etapa de análise de requisitos teve como finalidade identificar as necessidades funcionais e não funcionais do sistema, assegurando que o desenvolvimento atendesse às expectativas dos usuários e aos objetivos definidos.

Foram considerados como usuários-alvo os gestores acadêmicos, coordenadores de curso e pesquisadores da universidade, que necessitam de uma ferramenta de consulta e análise de indicadores de forma ágil e confiável.

A análise foi conduzida com base nas boas práticas de engenharia de software, utilizando levantamento de requisitos a partir de observação e estudo dos processos de gestão universitária existentes. A seguir, são descritos os requisitos funcionais e não funcionais identificados.

4.3.1. Requisitos Funcionais

Os requisitos funcionais definem as funções que o sistema deve executar para atender às suas finalidades. Entre os principais requisitos identificados estão:

- O sistema deve importar e processar dados dos arquivos *base.xlsx* e *formandos.xlsx*;
- Deve permitir a filtragem interativa de informações por curso, ano e tipo de convocação;
- Deve gerar gráficos dinâmicos e indicadores-chave (KPIs) com atualização automática;
- Deve possibilitar o cruzamento entre ingressantes e formandos, identificando padrões temporais;
- Deve incluir um módulo de previsões estatísticas, aplicando modelos Naive, Média Móvel, Holt e Prophet;
- Deve oferecer a opção de download dos dados filtrados em formato .csv;
- Deve apresentar interface intuitiva e de fácil navegação para diferentes perfis de usuário.

Essas funcionalidades garantem que o sistema cumpra seu papel como ferramenta de apoio à gestão acadêmica, combinando praticidade, desempenho e transparência.

4.3.2. Requisitos Não-Funcionais

Os requisitos não funcionais referem-se a características de qualidade do sistema, não relacionadas diretamente à execução de funções específicas:

- **Desempenho:** o sistema deve carregar e renderizar visualizações em tempo inferior a cinco segundos;
- **Usabilidade:** a interface deve ser clara, organizada e responsiva;
- **Portabilidade:** o aplicativo deve funcionar em diferentes sistemas operacionais (Windows, Linux, macOS);
- **Escalabilidade:** deve permitir a inclusão de novos cursos, anos e indicadores sem alteração do código principal;
- **Reprodutibilidade:** o código deve manter compatibilidade com versões futuras das bibliotecas utilizadas;
- **Manutenibilidade:** o código deve ser modular, com funções bem documentadas e reaproveitáveis;
- **Segurança:** os arquivos locais devem ser tratados sem envio de dados sensíveis a servidores externos.

Esses critérios garantem que o sistema apresente confiabilidade, desempenho e longevidade, sendo adequado para uso institucional e acadêmico.

4.4. Base de Dados

A base de dados utilizada neste trabalho foi composta por planilhas estruturadas em formato *Excel* (.xlsx), consolidadas a partir dos relatórios extraídos do Anuário Estatístico da UNESP e das Estatísticas da VUNESP.

Optou-se por armazenar os dados localmente, devido à simplicidade e flexibilidade do formato, permitindo leitura direta pelas bibliotecas *pandas* e *openpyxl* sem necessidade de um banco de dados relacional.

No entanto, o sistema foi projetado de forma escalável e compatível com ambientes em nuvem, permitindo que, futuramente, as planilhas possam ser hospedadas em serviços como Google Drive, SharePoint ou AWS S3, possibilitando

manutenção colaborativa por parte da universidade.

Essa estrutura híbrida — local e potencialmente online — facilita a atualização periódica dos dados pela diretoria acadêmica e reduz a necessidade de suporte técnico especializado.

Dessa forma, a base de dados cumpre papel central no sistema, servindo como fonte confiável e flexível para a geração dos indicadores e previsões acadêmicas apresentados no painel.

4.5. Estrutura e Arquitetura do Código

O sistema de visualização de indicadores acadêmicos foi concebido com o objetivo de transformar dados dispersos e estáticos em informações interativas, acessíveis e de fácil interpretação para gestores, docentes e pesquisadores. Todo o desenvolvimento foi orientado pela premissa de que uma ferramenta tecnológica, além de eficiente, deve ser compreensível e adaptável à realidade institucional, podendo ser mantida e expandida por diferentes usuários ao longo do tempo.

O código-fonte principal está concentrado no arquivo *app.py*, desenvolvido em Python, e estruturado de forma modular, legível e documentada. A arquitetura segue uma lógica linear de execução — carregamento → processamento → análise → visualização —, garantindo organização, rastreabilidade e manutenção simplificada.

O sistema foi desenvolvido e validado utilizando dados reais provenientes do Anuário Estatístico da UNESP e das estatísticas de vestibular da VUNESP, abrangendo diferentes cursos de graduação e anos de referência. O resultado é uma aplicação web interativa, leve e escalável, que permite exploração analítica em tempo real, sem a necessidade de conhecimento técnico em programação.

4.6. Carregamento e Tratamento dos Dados

A etapa inicial é responsável pelo carregamento, validação e preparação dos dados, etapa fundamental para assegurar a consistência e a confiabilidade das análises.

A leitura é realizada a partir de arquivos *.x/sx*, utilizando a biblioteca *pandas*, amplamente empregada em aplicações de ciência de dados. Durante o processo, o

sistema converte automaticamente os tipos de dados, remove inconsistências e normaliza campos textuais, como nomes de cursos e modalidades de ingresso, garantindo padronização entre os diferentes anos e fontes.

A função interna *load_data()* é responsável pela importação e tratamento dos dados. É decorada com cache para otimizar o desempenho, evitando recarregamentos desnecessários. Essa função também gera métricas derivadas, como taxa de ocupação de vagas e percentual de mulheres matriculadas, enriquecendo as bases antes das etapas de cálculo e visualização.

Esse tratamento prévio garante integridade e consistência estatística, minimizando erros de leitura e assegurando que as análises e gráficos subsequentes reflitam fielmente os dados institucionais.

4.7. Geração dos Indicadores e Métricas

Com as bases tratadas, o sistema calcula indicadores institucionais em tempo real, por meio de funções analíticas que agregam e cruzam os dados segundo curso, gênero, tipo de ingresso e período.

Esses indicadores são exibidos por meio de gráficos interativos desenvolvidos com a biblioteca Altair, que possibilita visualizações vetoriais limpas, escaláveis e altamente responsivas. Entre os principais indicadores calculados estão:

- Total de matrículas por curso e ano;
- Total de formandos por curso e ano;
- Distribuição percentual de alunos por gênero;
- Evolução temporal das matrículas e formaturas;
- Variação anual e taxas de crescimento ou retração;
- Cruzamento de dados de ingressantes e formandos por curso e/ou por ano ;
- Distribuição do método de ingresso por curso e/ou por ano evidenciando o programa de cotas.

Cada gráfico foi projetado com foco na clareza visual e interpretabilidade, utilizando cores neutras, legendas descritivas e escalas consistentes. A atualização ocorre de forma dinâmica, refletindo instantaneamente qualquer filtro aplicado pelo usuário.

4.8. Interface do Dashboard (Streamlit)

A interface gráfica, desenvolvida com Streamlit, foi desenhada com ênfase na usabilidade e acessibilidade. O painel principal adota uma estrutura de abas temáticas, dispostas horizontalmente no topo da tela, enquanto os filtros interativos são posicionados na barra lateral esquerda, permitindo uma navegação fluida e intuitiva.

Essa organização segue boas práticas de design de dashboards, priorizando a legibilidade e reduzindo a carga cognitiva do usuário. Cada aba corresponde a um módulo analítico independente, mas interconectado, o que reforça o conceito de análise integrada.

O sistema foi otimizado para responder em tempo real às interações do usuário, qualquer ajuste em filtros ou seleções de curso, ano ou modalidade de ingresso atualiza imediatamente os gráficos e tabelas exibidas.

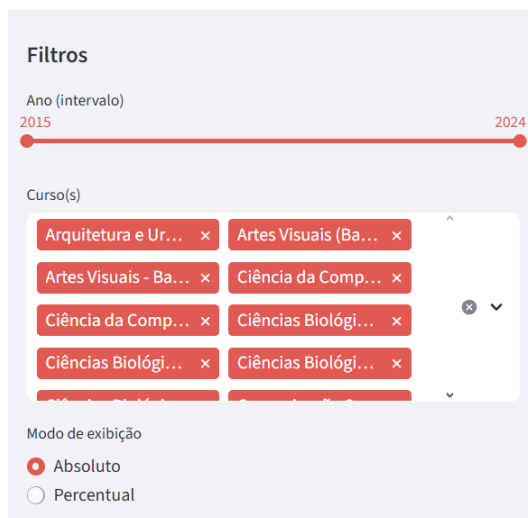
4.9. Menu lateral e filtros interativos

O menu lateral do sistema (Figura 3) concentra os principais filtros analíticos, que permitem selecionar períodos específicos (anos), cursos, gêneros ou modalidades de ingresso.

Esses filtros são aplicados dinamicamente, sem a necessidade de recarregar o painel. A arquitetura de cache inteligente do Streamlit garante desempenho fluido mesmo com bases de dados extensas, mantendo o tempo médio de resposta inferior a um segundo por interação.

Essa abordagem oferece ao usuário autonomia analítica, permitindo explorar diferentes cenários e recortes sem conhecimentos técnicos de programação ou banco de dados.

Figura 3 - Aplicativo: Filtros Interativos



Fonte: Elaborado pelo Autor.

4.10. Abas Principais do Sistema

O sistema está dividido em nove abas principais (Figura 4), cada uma voltada a uma dimensão específica da análise institucional. Essa segmentação foi pensada para reproduzir o fluxo natural de investigação de um gestor ou pesquisador acadêmico.

Figura 4 - Aplicativo: Abas

Visão Geral Por Gênero Por Curso Insights Formandos Cruzado Previsões Programa de Cotas

Fonte: Elaborado pelo Autor.

4.10.1. Aba “Visão Geral”

Apresenta um panorama consolidado das matrículas ao longo dos anos (Figura 6), exibindo gráficos de evolução e composição por tipo de convocação. São destacados indicadores-chave de desempenho (KPIs), como total de matrículas, variação em relação ao ano anterior, percentual de estudantes do sexo feminino e tipo de ingresso predominante (Figura 5).

Figura 5 - Aplicativo: Visão Geral Cabeçalho

Sistema de Visualização de Indicadores Acadêmicos UNESP - Bauru

Total de matrículas (filtro)

13.118

Variação vs. ano anterior

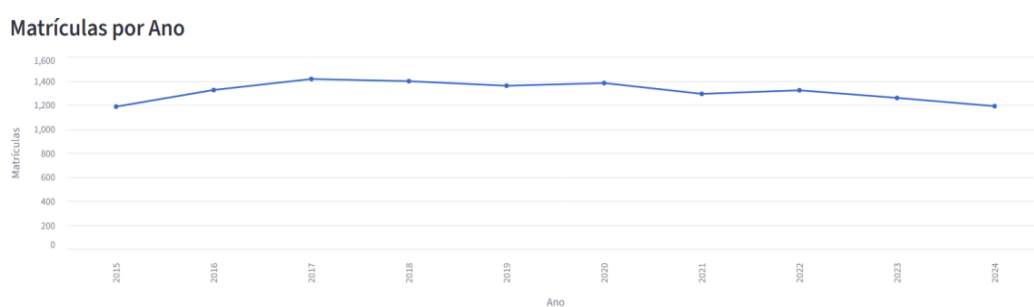
-5.4%

% Feminino (último ano)

49.4%

Fonte: Elaborado pelo Autor.

Figura 6 - Aplicativo: Matrículas por Ano

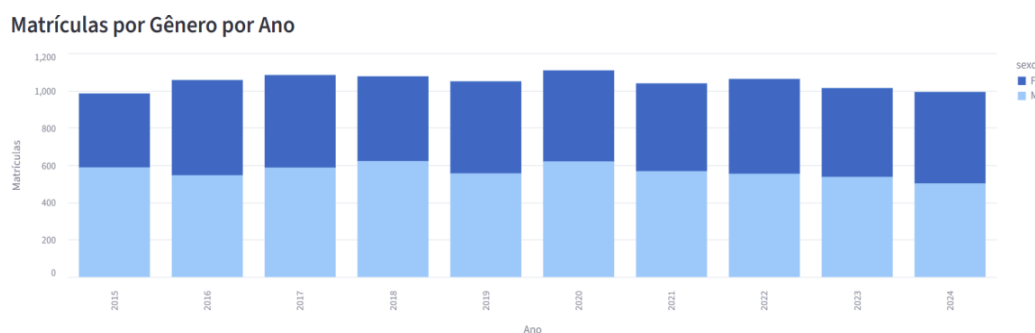


Fonte: Elaborado pelo Autor.

4.10.2. Aba "Por Gênero"

Exibe a distribuição das matrículas entre homens e mulheres (Figura 7), permitindo identificar tendências de equidade de gênero e acompanhar a evolução da participação feminina nos cursos de graduação.

Figura 7 - Aplicativo: Aba "Por Gênero"

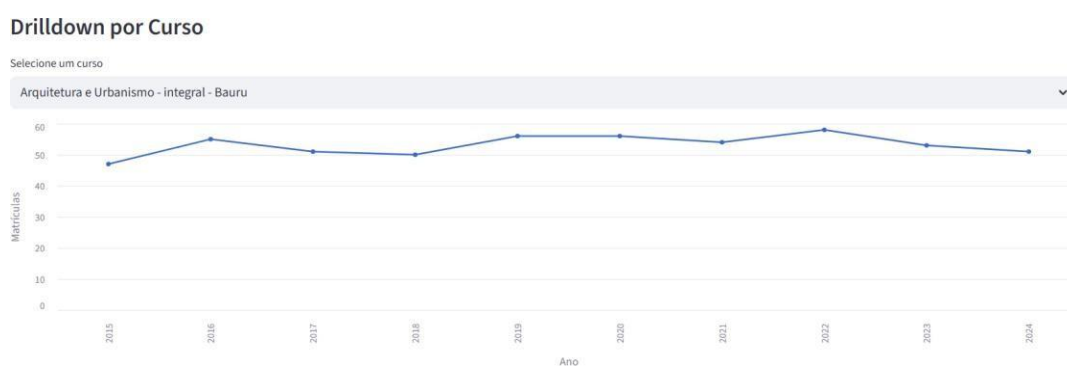


Fonte: Elaborado pelo Autor.

4.10.3. Aba “Por Curso”

Permite analisar detalhadamente cada curso (Figura 8), com evolução de matrículas, séries históricas específicas e variações interanuais. Cada visualização oferece botões de exportação em .csv, incentivando o uso dos dados em pesquisas e relatórios externos.

Figura 8 - Aplicativo: Drilldown por Curso

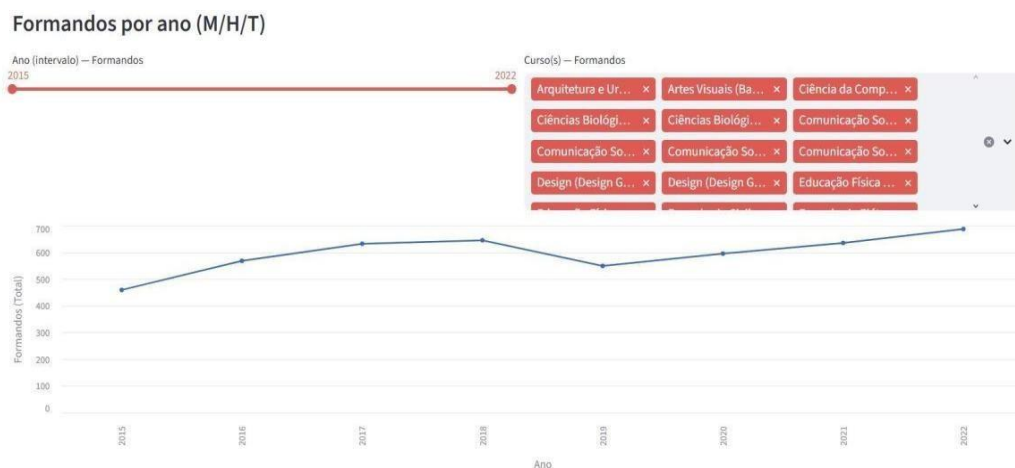


Fonte: Elaborado pelo Autor.

4.10.4. Aba “Formandos”

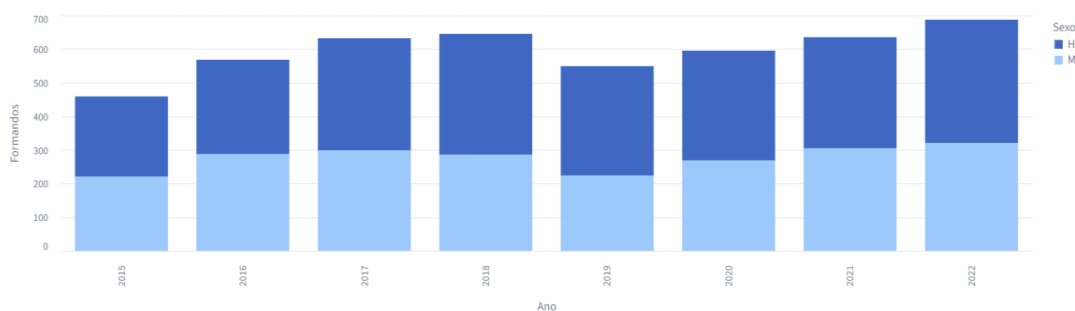
Centraliza os dados referentes aos concluintes de cada curso (Figura 11), permitindo observar a evolução das formaturas (Figura 9), a distribuição por gênero e a taxa média de crescimento anual (Figura 10). A leitura é feita diretamente do arquivo formandos.xlsx, tratado com a mesma padronização dos dados de ingressantes.

Figura 9 - Aplicativo: Formandos por Ano



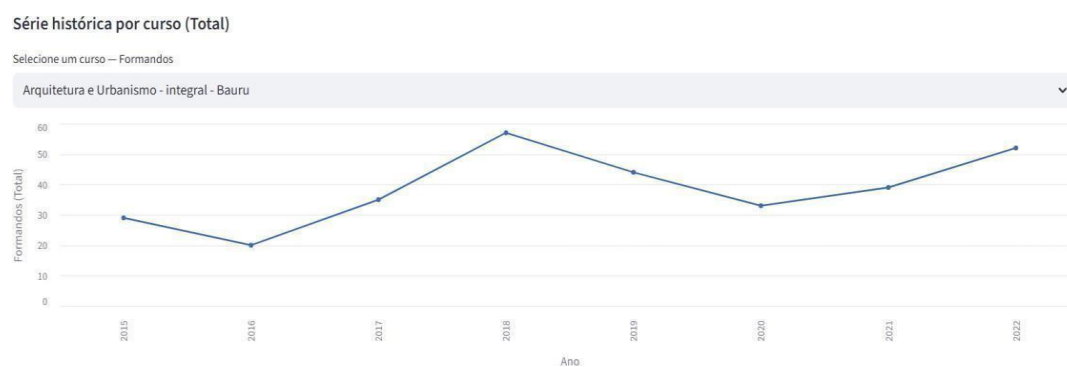
Fonte: Elaborado pelo Autor.

Figura 10 - Aplicativo: Formandos por Gênero



Fonte: Elaborado pelo Autor.

Figura 11 - Aplicativo: Série Histórica por Curso



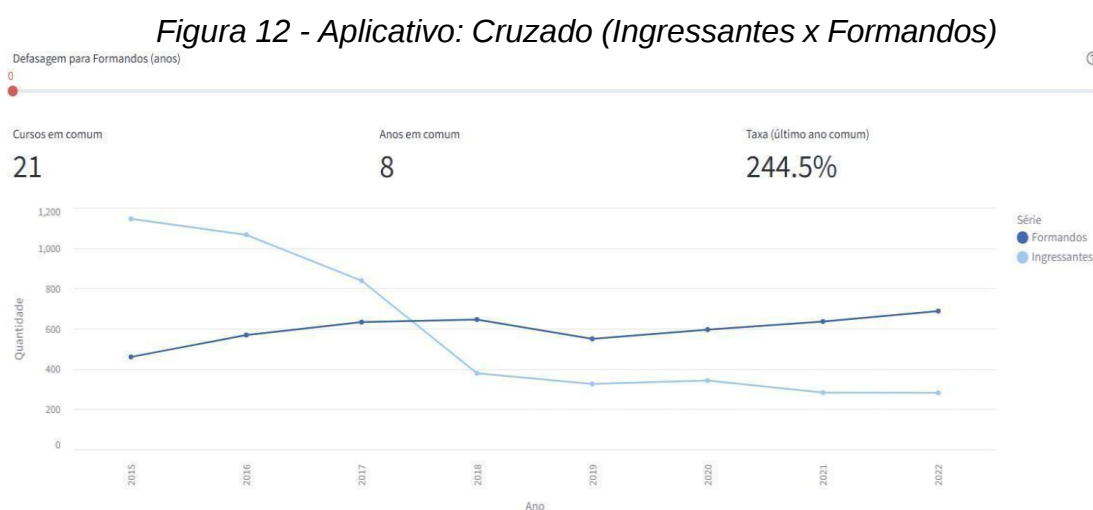
Fonte: Elaborado pelo Autor.

4.10.5. Aba “Cruzado: Ingressantes x Formandos”

Realiza o cruzamento entre ingressantes e formandos considerando apenas os cursos comuns a ambas as bases (Figura 12). Permite observar a relação temporal entre o número de matrículas e o de formaturas, evidenciando taxas de conclusão e possíveis defasagens entre entrada e saída.

Durante o desenvolvimento do módulo “Cruzado: Ingressantes x Formandos”, identificou-se uma limitação importante relacionada à correspondência temporal entre os alunos ingressantes e os formandos. Os dados institucionais disponibilizados pelo *Anuário Estatístico da UNESP* e pela *VUNESP* não permitem rastrear individualmente o percurso de cada estudante, impossibilitando determinar com exatidão quais formandos pertencem a cada coorte de ingresso. Assim, ao comparar diretamente o número de ingressantes e formandos por ano, sem qualquer ajuste, corre-se o risco de associar erroneamente turmas distintas, distorcendo a interpretação dos indicadores de conclusão.

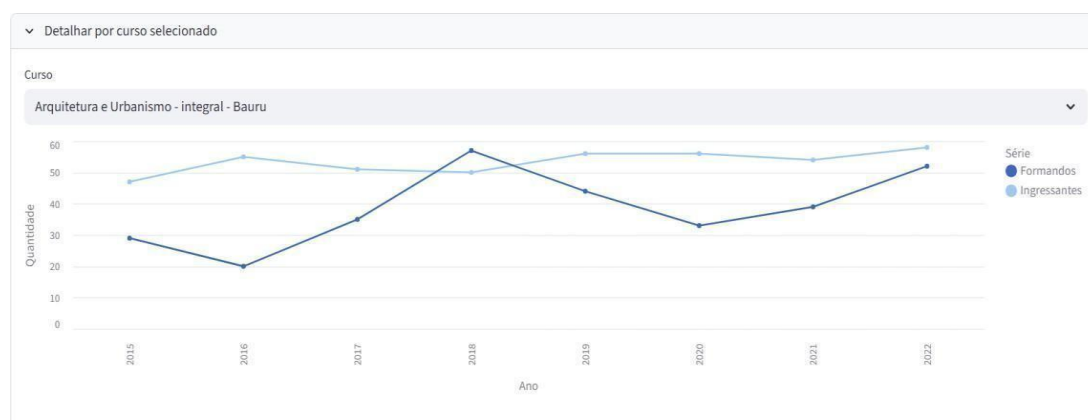
Para mitigar esse problema, foi implementado no sistema um mecanismo de defasagem temporal, representado por um controle deslizante (“offset”) que permite deslocar os anos de formatura em relação aos anos de ingresso. Esse recurso aplica a operação $(ano_formando - defasagem)$, ajustando a série temporal dos formandos de modo a alinhar o momento esperado de conclusão com o ano correspondente de ingresso.



Fonte: Elaborado pelo Autor.

Dessa forma, quando a defasagem é configurada, por exemplo, em quatro anos, o sistema compara os formandos de 2023 com os ingressantes de 2019, simulando o ciclo típico de um curso de quatro anos de duração. Esse alinhamento permite uma análise mais realista das taxas de conclusão e dos fluxos acadêmicos, aproximando o comportamento estatístico do sistema ao percurso médio dos estudantes.

Figura 13 - Aplicativo: Cruzado por Curso



Fonte: Elaborado pelo Autor.

A ausência de tal ajuste levaria a análises inconsistentes, como taxas de conclusão subestimadas ou superestimadas, especialmente em cursos com durações distintas. Assim, a implementação da defasagem torna possível representar, de maneira aproximada, a dinâmica longitudinal entre ingresso e formatura, ainda que os dados disponíveis sejam agregados e não individualizados, ampliando significativamente a interpretabilidade dos resultados (Figura 13).

4.10.6. Aba “Previsões”

Aplica modelos Naive, Média Móvel, Holt e Prophet às séries temporais de ingressantes e formandos, gerando projeções de curto e médio prazo (Figura 14). Os resultados são exibidos em gráficos com as curvas de tendência histórica e a faixa de confiança das previsões. Esse módulo reforça o caráter analítico do sistema, oferecendo um componente preditivo para planejamento institucional.

Figura 14 - Aplicativo: Previsões



Fonte: Elaborado pelo Autor.

4.10.7. Aba “Programa de Cotas”

Apresenta uma análise detalhada das modalidades de ingresso, comparando os resultados entre SRVEBP, SU e PPI (Figura 15). O painel mostra a evolução da representatividade das cotas ao longo dos anos e permite avaliar o impacto das políticas de inclusão na universidade.

É possível avaliar uma visualização de série histórica para cada curso selecionado de determinado período filtrado (Figura 16).

Figura 15 - Aplicativo: Programa de Cotas



Fonte: Elaborado pelo Autor.

Figura 16 - Aplicativo: Série Histórica Programa de Cotas por Curso



Fonte: Elaborado pelo Autor.

4.10.8. Exportação e Atualização dos Dados

O sistema disponibiliza botões para download dos dados filtrados em formato .csv, permitindo que os usuários realizem análises adicionais fora da plataforma em todas as abas e módulos analíticos. Cada aba apresenta também uma pré-visualização do consolidado dos dados correspondentes, possibilitando ao usuário verificar as informações que serão exportadas antes de realizar o *download* (Figura 17).

Como o Streamlit lê diretamente os arquivos .xlsx atualizados, a simples

substituição dos arquivos de entrada é suficiente para que todas as análises e gráficos sejam atualizados automaticamente, sem necessidade de modificações no código-fonte. Essa arquitetura foi planejada para garantir facilidade de manutenção e escalabilidade futura do projeto, de modo que a atualização dos dados possa ser realizada apenas pela inserção de um novo ano na base de dados existente, sem demandar ajustes técnicos adicionais.

Figura 17 - Aplicativo: Exportação de Dados

Tabela — Dados de cotas filtrados

	curso	vagas	SU	SRVEBP	ppi	ano
8	Arquitetura e Urbanismo - integral - Bauru	45	33	8	4	2015
0	Ciência da Computação (Bac) - integral - Bauru	30	22	5	3	2015
9	Design (Design Gráfico ou de Produto) - noturno - Bauru	60	45	10	5	2015
11	Design (Design Gráfico ou de Produto) - noturno - Bauru	60	45	10	5	2015
10	Design (Design Gráfico) - diurno - Bauru	30	22	5	3	2015
12	Design (Design Gráfico) - diurno - Bauru	30	22	5	3	2015
1	Engenharia Civil - integral - Bauru	60	45	10	5	2015
2	Engenharia Elétrica - integral - Bauru	60	45	10	5	2015
3	Engenharia Mecânica - integral - Bauru	60	45	10	5	2015
4	Física (Lic - noturno) / (Bac em Física de Materiais - vespertino/noturno) - Ba	60	45	10	5	2015

Baixar dados de cotas (CSV)

Fonte: Elaborado pelo Autor.

4.11. Módulo de Previsões e Análises Temporais

O módulo de previsão foi desenvolvido para complementar as análises históricas com projeções futuras, permitindo estimar o comportamento esperado dos indicadores de ingressantes e formandos. Foram implementados os métodos Naive, Média Móvel, Holt e Prophet, cada um com características e níveis de complexidade distintos.

O método Naive assume que o valor futuro será igual ao último valor observado, servindo como linha de base (*baseline*) para comparação de desempenho. A Média Móvel, por sua vez, calcula a média dos três últimos valores históricos e utiliza esse resultado como previsão para os períodos seguintes, suavizando flutuações pontuais. O modelo Holt, pertencente à família de suavização exponencial, considera o nível e a tendência linear da série, sendo capaz de capturar movimentos de crescimento ou

queda de forma mais realista. Por fim, o modelo Prophet, desenvolvido pela Meta (Facebook), aplica decomposição aditiva para modelar tendência e sazonalidade de maneira automática, apresentando também intervalos de confiança que refletem a incerteza associada às estimativas. As curvas projetadas mostraram coerência com o comportamento histórico, especialmente para cursos com séries temporais longas e regulares, demonstrando a viabilidade do uso dessas técnicas para planejamento acadêmico (Figura 14).

As previsões foram geradas para horizontes de até três anos à frente, permitindo avaliar possíveis cenários de evolução das matrículas. A avaliação de desempenho foi realizada com base em métricas como o erro quadrático médio (RMSE) e o erro percentual médio absoluto (MAPE), utilizando validação *holdout* dos dois últimos anos da série histórica.

As projeções resultantes apresentaram coerência com os padrões observados nas séries históricas, especialmente em cursos com registros regulares e consistentes, demonstrando a viabilidade do uso combinado desses métodos como instrumento de apoio ao planejamento e à tomada de decisão no contexto da gestão acadêmica.

Para cada previsão gerada, o sistema apresenta três indicadores principais que auxiliam na interpretação da qualidade e da confiabilidade dos resultados: Pontos históricos, RMSE (*holdout* 2 anos) e MAPE (*holdout* 2 anos).

O indicador Pontos históricos representa a quantidade de observações disponíveis na série temporal utilizada para o treinamento do modelo. No exemplo exibido, há sete pontos históricos, o que corresponde a sete anos consecutivos de dados de matrículas para o curso selecionado. Esse número define a base sobre a qual os modelos de previsão foram ajustados — quanto maior a série, maior tende a ser a estabilidade estatística das estimativas.

O RMSE, ou raiz do erro quadrático médio, mede a diferença média entre os valores previstos e os valores observados, penalizando erros maiores de forma mais intensa. No caso apresentado, o RMSE de 4,7 indica que, em média, as previsões diferem cerca de 4,7 matrículas em relação aos dados reais dos dois últimos anos da série (*holdout*). Valores menores de RMSE indicam maior precisão do modelo.

O MAPE, ou erro percentual médio absoluto, expressa a diferença média entre as previsões e os valores reais em termos percentuais. Um MAPE de 13,7% significa

que, em média, as previsões apresentam erro de aproximadamente 13,7% em relação aos valores observados. Essa métrica é amplamente utilizada em modelos preditivos por permitir uma comparação intuitiva entre cursos e séries de diferentes magnitudes.

Essas métricas são calculadas a partir de uma validação do tipo *holdout*, que reserva os dois últimos anos da série para teste, utilizando os dados anteriores para o treinamento do modelo (Figura 18). Essa abordagem permite avaliar a capacidade preditiva real do método escolhido, oferecendo uma estimativa prática da sua precisão e confiabilidade antes de projetar valores para anos futuros.

Figura 18: RMSE e MAPE

Pontos históricos	RMSE (holdout 2 anos)	MAPE (holdout 2 anos)
10	6.1	11.6%

Fonte: Elaborada pelo Autor.

4.12. Avaliação de desempenho e usabilidade do sistema

O desempenho do sistema foi avaliado quanto ao tempo de carregamento, responsividade das interações e clareza das visualizações. Em testes realizados no hardware de desenvolvimento (Samsung Galaxy Book3 360, Intel Core i5, 16 GB RAM), o tempo médio de inicialização do painel foi inferior a 5 segundos, e a renderização dos gráficos ocorreu de forma fluida.

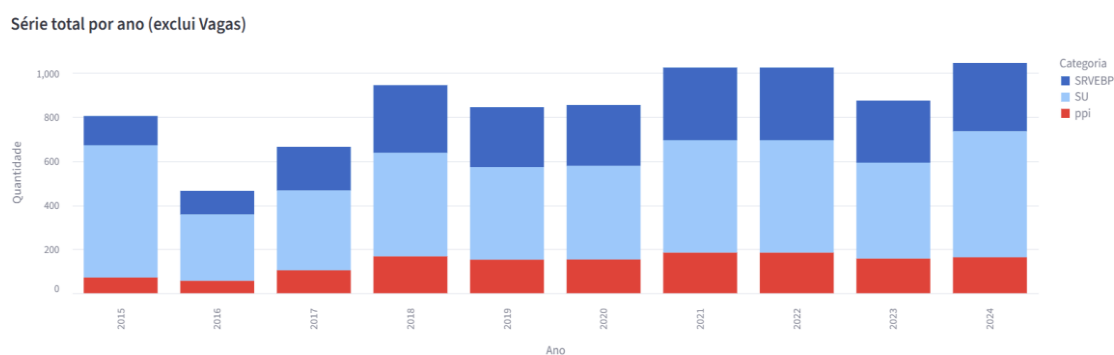
A usabilidade foi avaliada de modo exploratório, priorizando a intuição da navegação e a facilidade de filtragem. A disposição das abas e o uso de cores discretas, em contraste com o fundo escuro, contribuem para uma leitura confortável e uma experiência agradável ao usuário.

4.13. Estudo de Caso: Evolução do Programa de Cotas na UNESP

A análise apresentada nesta seção, exemplifica uma das aplicações práticas do sistema desenvolvido, demonstrando sua capacidade de sintetizar dados

institucionais e gerar insights sobre políticas públicas de acesso e inclusão. Utilizando a aba “Programa de Cotas”, é possível observar de forma clara e interativa a evolução da participação das diferentes modalidades de ingresso ao longo da última década.

Figura 19 - Aplicativo: Programa de Cotas Caso de Estudo



Fonte: Elaborado pelo Autor.

A Figura 19 apresenta a série total por ano (2015–2024), excluindo o número de vagas disponíveis e considerando apenas os ingressantes efetivos, distribuídos segundo as categorias SRVEBP (Escola Pública com Reserva de Vaga e Benefício de PPI e/ou Renda), SU (Escola Pública Sem Benefício) e PPI.

O gráfico evidencia o crescimento progressivo da representatividade das políticas de cotas dentro da universidade. Enquanto em 2015 a participação das categorias vinculadas ao programa era significativamente menor, a partir de 2018 nota-se uma expansão contínua e consolidada, com destaque para o aumento do grupo SRVEBP.

Esses resultados estão em consonância com os dados divulgados no relatório “Renda e Unidade Universitária 2014–2024” (UNESP, 2024), que aponta para uma mudança estrutural no perfil socioeconômico dos ingressantes, reforçando o papel da política de reserva de vagas na ampliação do acesso de estudantes de baixa renda e oriundos de escolas públicas. O documento destaca que, em diversas unidades da universidade, mais de 50% dos novos estudantes passaram a ingressar por meio das cotas após a implementação do sistema de ampla reserva de vagas.

De forma complementar, o “Relatório COPE 2024” reforça que o avanço observado nas estatísticas de ingresso não apenas confirma a efetividade da política, mas também indica melhor desempenho e permanência dos alunos cotistas,

demonstrando que a ampliação do acesso veio acompanhada de melhorias nos indicadores de conclusão e redução da evasão.

A ferramenta desenvolvida neste trabalho permite que essas tendências sejam visualizadas e quantificadas em tempo real, algo que antes exigia análise manual de planilhas extensas e relatórios estáticos. Por meio do painel interativo, o usuário pode selecionar intervalos específicos, comparar categorias e identificar, visualmente, a evolução do programa de cotas ao longo do tempo.

Essa capacidade de integração e análise dinâmica transforma o painel em uma ferramenta estratégica de monitoramento institucional, oferecendo suporte técnico à Comissão Permanente de Políticas de Equidade (COPE), aos gestores universitários e aos pesquisadores interessados na avaliação das políticas de inclusão.

Assim, a análise da Figura 19 demonstra como o sistema contribui diretamente para o fortalecimento da cultura de gestão baseada em dados na UNESP, permitindo que decisões sejam fundamentadas em informações atualizadas, acessíveis e visualmente claras. A evolução visível das cotas entre 2015 e 2024 reforça não apenas o impacto das políticas implementadas, mas também a utilidade prática e a relevância científica da ferramenta desenvolvida.

4.14. Discussão dos resultados obtidos

Os resultados obtidos com o desenvolvimento do sistema confirmam sua eficácia como ferramenta de apoio à gestão acadêmica. O painel consolidou, em um único ambiente interativo, informações anteriormente dispersas em relatórios e planilhas, permitindo análises históricas e preditivas de forma integrada e acessível.

A aplicação dos métodos de previsão Naive, Média Móvel, Holt e Prophet demonstrou que é possível gerar estimativas coerentes e informativas mesmo a partir de séries históricas curtas, fornecendo subsídios quantitativos para o planejamento institucional. Esses modelos, de diferentes níveis de complexidade, permitiram comparar abordagens simples e avançadas, contribuindo para a compreensão da tendência de evolução dos indicadores acadêmicos. Além disso, a utilização de tecnologias livres e o formato acessível do dashboard favorecem a reprodutibilidade e a escalabilidade do sistema, que pode ser facilmente adaptado para outras unidades da UNESP ou instituições de ensino com estrutura de dados semelhante.

Além disso, a utilização de tecnologias livres e de código aberto, como *Python*, *Streamlit* e *Altair*, aliada à estrutura modular do sistema, favorece sua reprodutibilidade e escalabilidade. O painel pode ser facilmente adaptado para outras unidades da UNESP ou replicado em instituições de ensino com estruturas de dados semelhantes, bastando a atualização dos arquivos de entrada para incluir novos períodos ou cursos.

Conclui-se que a solução desenvolvida atingiu plenamente os objetivos propostos, oferecendo um ambiente de análise interativo, transparente e cientificamente fundamentado, capaz de fortalecer a cultura de gestão universitária baseada em dados e de apoiar a tomada de decisões estratégicas no contexto acadêmico.

5. CONCLUSÃO

O desenvolvimento do sistema de visualização de indicadores acadêmicos alcançou plenamente o objetivo de oferecer uma ferramenta prática e eficaz para apoiar a gestão universitária no campus da UNESP Bauru. Partindo da necessidade identificada na introdução — a ausência de instrumentos integrados para análise e comunicação de dados acadêmicos —, o projeto consolidou informações antes dispersas entre relatórios e planilhas, transformando-as em um painel interativo capaz de gerar análises dinâmicas e de fácil interpretação.

A integração de dados provenientes do Anuário Estatístico da UNESP e das estatísticas da VUNESP permitiu estruturar uma base unificada e confiável, sobre a qual foram calculados indicadores de matrícula, conclusão, distribuição por gênero e modalidade de ingresso. O sistema demonstrou ser eficiente tanto no processamento quanto na visualização das informações, utilizando Python e bibliotecas de código aberto como pandas, NumPy, Altair e Streamlit, que se mostraram adequadas para a criação de uma aplicação leve, acessível e escalável.

A inclusão de um módulo de previsão temporal, baseado nos métodos Naive, Média Móvel, Holt e Prophet, ampliou a dimensão analítica da ferramenta, possibilitando a estimativa de tendências futuras de ingresso e formatura. As previsões obtidas apresentaram coerência com os dados históricos, reforçando o potencial do sistema para apoiar o planejamento institucional e a avaliação de políticas de acesso e permanência.

Do ponto de vista técnico, a arquitetura modular e o uso de bibliotecas amplamente difundidas asseguram a reprodutibilidade, manutenção e expansão do sistema, que pode ser facilmente atualizado com novos anos ou novas bases de dados, sem necessidade de alterações no código-fonte. Do ponto de vista institucional, o painel contribui para a consolidação de uma cultura de gestão orientada por evidências, promovendo maior transparência, eficiência e autonomia nas decisões acadêmicas.

Conclui-se, portanto, que o sistema desenvolvido não apenas cumpre os objetivos propostos, mas também representa uma solução tecnológica inovadora e replicável, capaz de fortalecer a governança universitária e de servir como base para futuras aplicações de ciência de dados no contexto acadêmico. Como perspectivas de

continuidade, recomenda-se a integração com bases de dados institucionais em tempo real e o aprimoramento dos módulos analíticos, de modo a expandir o escopo das previsões e das análises de desempenho acadêmico.

REFERÊNCIAS

ALTAR, Jake VanderPlas. *Altair: Declarative Visualization in Python*. Journal of Open Source Software, v. 3, n. 32, p. 1057, 2018. DOI: 10.21105/joss.01057.

DRUCKER, Peter. *Management Challenges for the 21st Century*. New York: HarperBusiness, 1999.

FEW, Stephen. *Information Dashboard Design: The Effective Visual Communication of Data*. 2. ed. Burlingame, CA: Analytics Press, 2012.

FIALHO, Francisco Antônio Pereira; SILVA, Rafael Bianchi da. *Gestão Universitária Baseada em Evidências: Perspectivas e Desafios*. Revista Gestão Universitária, Florianópolis, v. 12, n. 2, 2020.

HOLT, Charles C. *Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages*. ONR Research Memorandum, Carnegie Institute of Technology, Pittsburgh, 1957.

HYNDMAN, Rob J.; ATHANASOPOULOS, George. *Forecasting: Principles and Practice*. 3. ed. Melbourne: OTexts, 2018. Disponível em: <https://otexts.com/fpp3/>. Acesso em: 8 out. 2025.

MCKINNEY, Wes. *Data Structures for Statistical Computing in Python*. In: *Proceedings of the 9th Python in Science Conference*, Austin, 2010.

OLIPHANT, Travis E. *A Guide to NumPy*. USA: Trelgol Publishing, 2006.

PENG, Roger D. *Reproducible Research in Computational Science*. Science, v. 334, n. 6060, p. 1226–1227, 2011. DOI: 10.1126/science.1213847.

STREAMLIT INC. *Streamlit Documentation*. 2023. Disponível em: <https://docs.streamlit.io/>. Acesso em: 8 out. 2025.

TAYLOR, Sean J.; LETHAM, Benjamin. *Forecasting at Scale*. The American Statistician, v. 72, n. 1, p. 37–45, 2018. DOI: 10.1080/00031305.2017.1380080.

TUFTE, Edward R. *The Visual Display of Quantitative Information*. 2. ed. Cheshire: Graphics Press, 2001.

UNESP – Universidade Estadual Paulista “Júlio de Mesquita Filho”. *Renda e Unidade Universitária 2014–2024*. Comissão Permanente de Políticas de Equidade (COPE), 2024. Disponível em: <https://www2.unesp.br/Home/cope/documentos/rendaunidadeuniv20142024.pdf>. Acesso em: 8 out. 2025.

UNESP – Universidade Estadual Paulista “Júlio de Mesquita Filho”. *Relatório COPE 2024*. Comissão Permanente de Políticas de Equidade (COPE), 2024. Disponível em: <https://www2.unesp.br/Home/cope/documentos/relatorio-cope->

[2024.pdf](#). Acesso em: 8 out. 2025.

UNESP – Universidade Estadual Paulista “Júlio de Mesquita Filho”. *Anuário Estatístico da UNESP*. São Paulo: Pró-Reitoria de Planejamento Estratégico e Gestão, 2025. Disponível em: <https://www2.unesp.br/portal#!/anuario>. Acesso em: 8 out. 2025.

VAN ROSSUM, Guido; DRAKE, Fred L. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.

VUNESP – Fundação para o Vestibular da Universidade Estadual Paulista. *Estatísticas de Vestibular*. São Paulo: Fundação VUNESP, 2025. Disponível em: <https://www.vunesp.com.br/institucional/estatisticavestibular>. Acesso em: 8 out. 2025