

## 1. INTRODUÇÃO

## 2. OBJETIVOS

### 2.1 Objetivos gerais

### 2.2 Objetivos específicos

#### 2.2.1 Cronograma do Projeto

## 3. JUSTIFICATIVA

## 4. REVISÃO BIBLIOGRÁFICA (SÍNTESE)

## 5. METODOLOGIA

### 5.1 Definição dos objetivos

### 5.2 Seleção das bases de dados

- Microdados do Censo da Educação Superior – INEP: são arquivos CSV volumosos

### 5.3 Pré-processamento e transformação

### 5.4 Mineração dos dados, EDA, Análises, MVP

### 5.5 Interpretação e avaliação dos resultados

## 6. RESULTADOS E DISCUSSÕES

## CRONOGRAMA DE ATIVIDADES (2025)

Definição do tema: Finalizado (13/03)

Coleta de dados INEP: Até 13/03

Pré-processamento e EDA: Até 13/03

Validação cruzada e gráficos: Até 16/03

Refinamento de análises: Até 22/03

Validação com orientadora: 24/03

Modelagem preditiva: 25/03 a 15/04

Validação modelo: 10/04 a 16/04

Otimização do modelo: 17/04 a 15/05

Segunda entrega: Até 06/05

Vídeo final: 06/05 a 20/05

Entrega final: Até 25/05

## **1. INTRODUÇÃO**

Este projeto tem como foco a análise de dados da educação superior brasileira, com ênfase na evasão e conclusão de cursos em instituições públicas e privadas.

## **3. JUSTIFICATIVA**

A evasão acadêmica impacta diretamente o desenvolvimento educacional e a alocação de recursos públicos. Compreender padrões e prever tendências é fundamental para a tomada de decisões mais eficazes por parte de gestores e formuladores de política.

## **5. METODOLOGIA**

### **5.1 Definição dos objetivos**

- Estimar taxas de ingresso, evasão e conclusão de cursos superiores no Brasil.
- Criar visualizações analíticas e um modelo preditivo para auxiliar diagnósticos.

### **5.2 Seleção das bases de dados**

- Utilização dos Microdados do Censo da Educação Superior – INEP (1995–2023).
- Arquivos CSV com milhões de registros, organizados anualmente.

### **5.3 Pré-processamento e transformação**

- Leitura com pandas; tratamento de separadores, codificações e colunas inconsistentes.
- Normalização com validações estatísticas e remoção de outliers.
- Scripts utilizados: ``pre_processamento.py``, ``tratar_dados.py``.

### **5.4 Mineração dos dados, EDA e MVP**

- Análises com ``analises.py``, incluindo correlações e geração de gráficos (Seaborn, Matplotlib).
- Scripts de visualização: ``gerar_graficos.py`` e ``gerar_relatorio_pdf.py``.

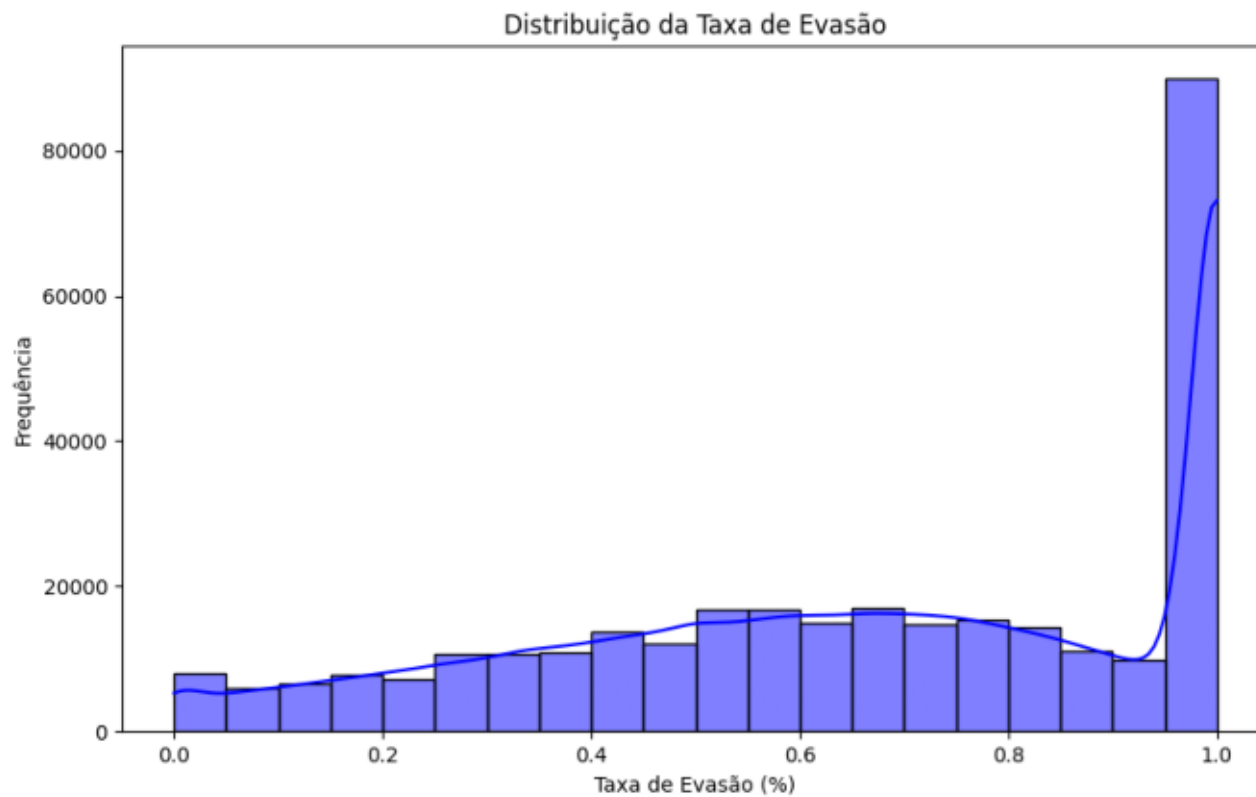
### **5.5 Modelagem e avaliação**

- Modelo preditivo de regressão linear implementado em ``treinamento_modelo.py``.
- Métricas avaliadas: erro quadrático médio e matriz de confusão.

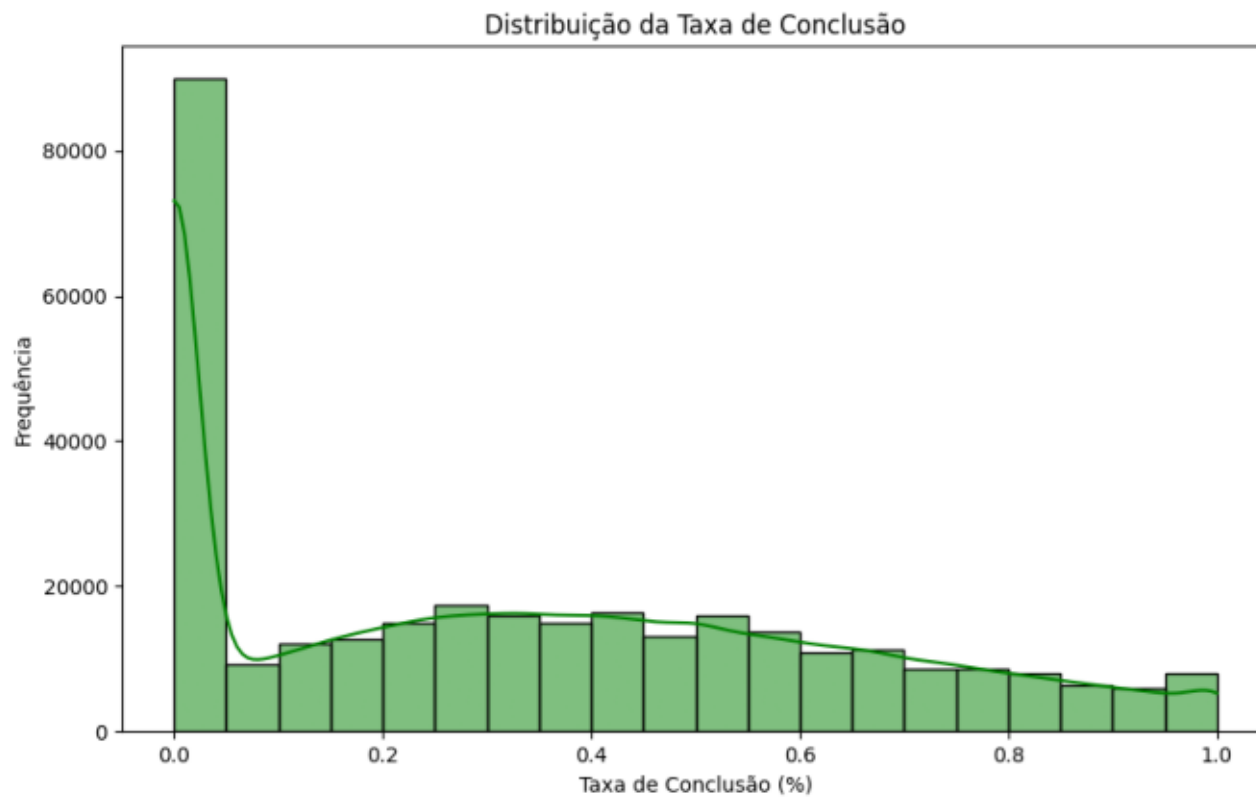
### **5.6 Tecnologias e ferramentas utilizadas**

- Linguagem: Python 3.11

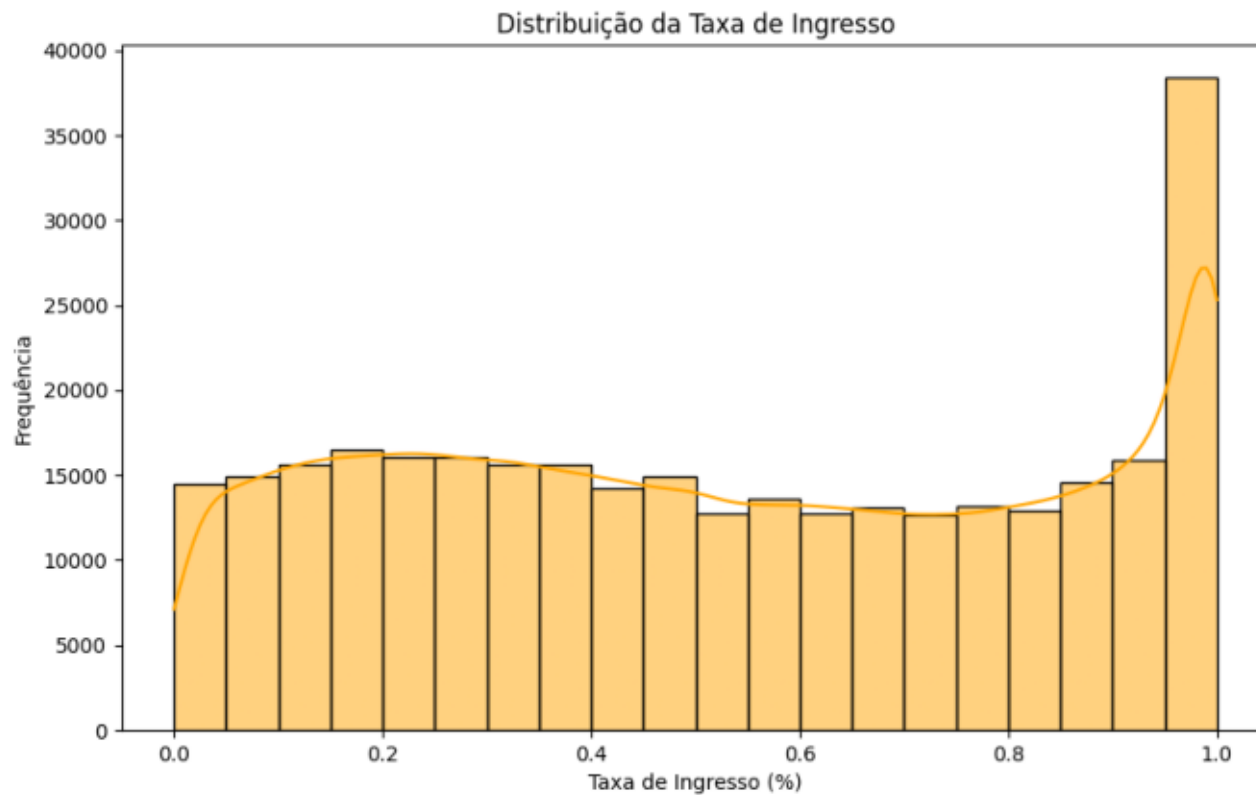
# Distribuicao Taxa Evasao



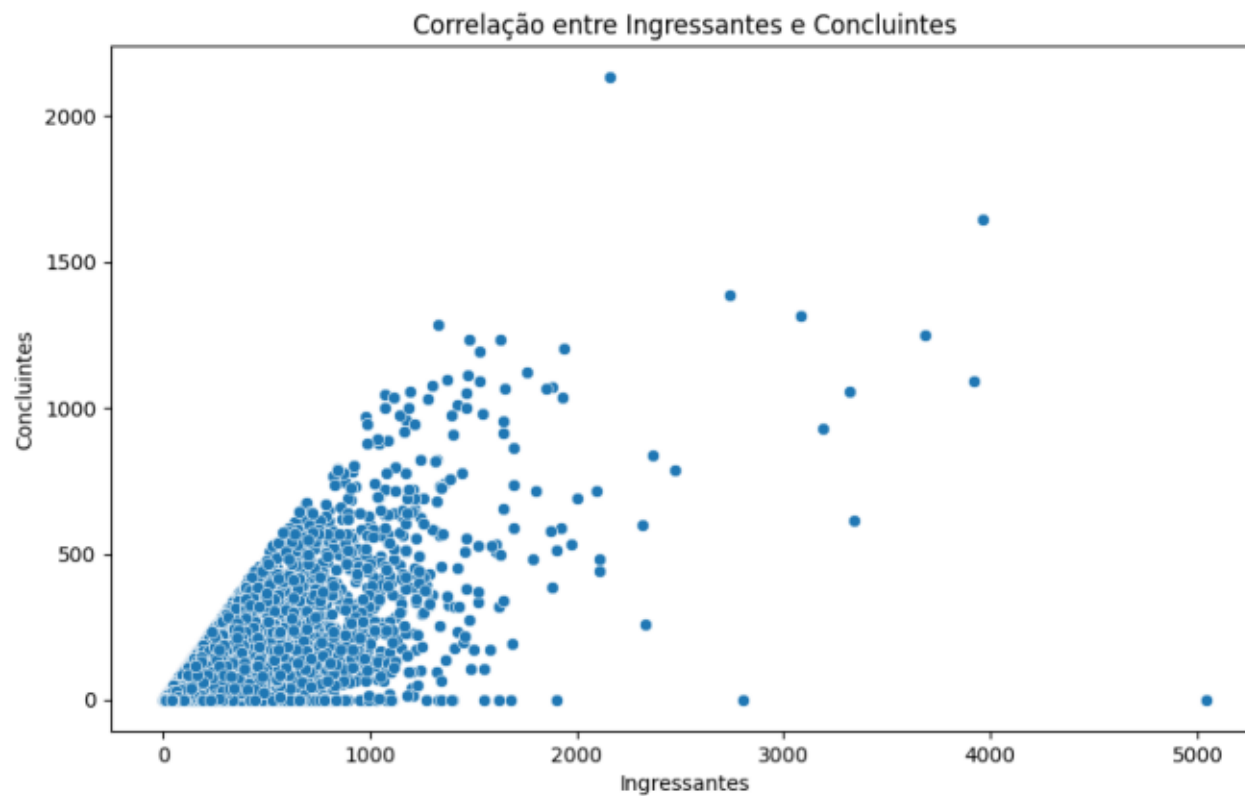
## Grafico Taxa Conclusao



## Grafico Taxa Ingresso

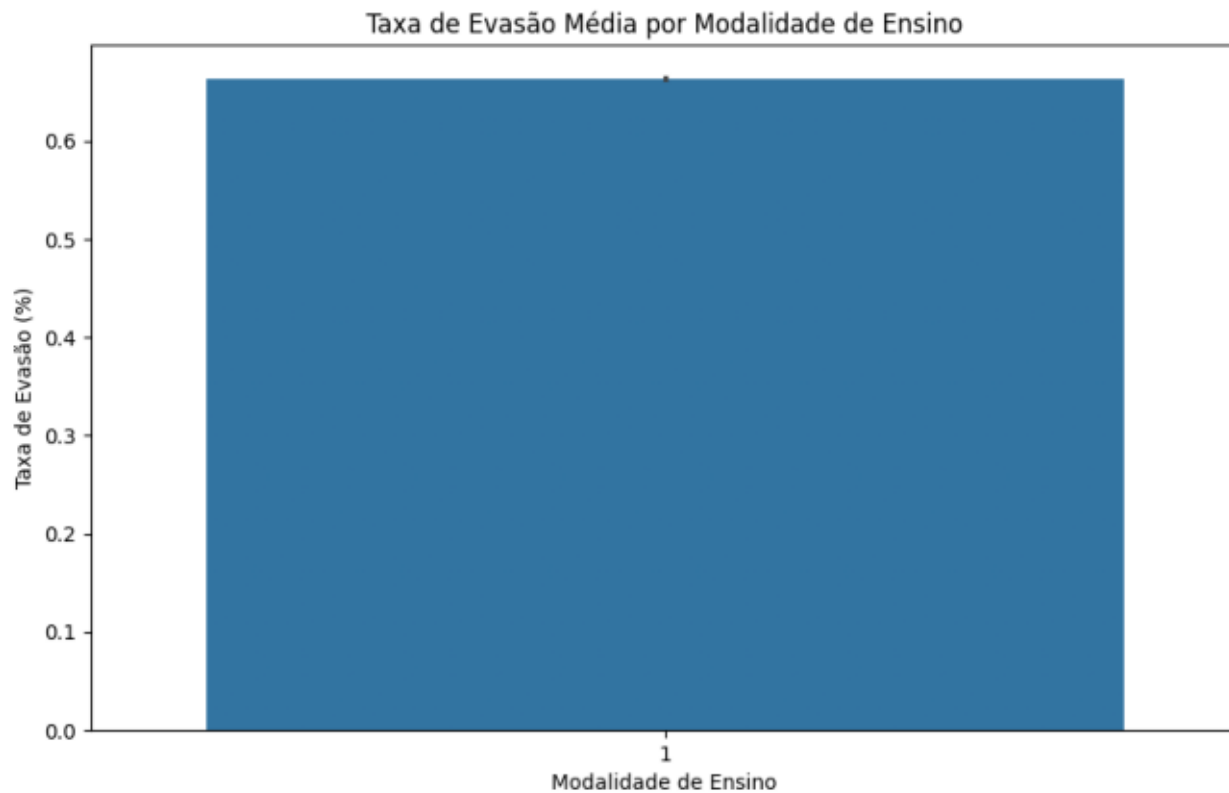


## Grafico Ingressantes Vs Concluintes

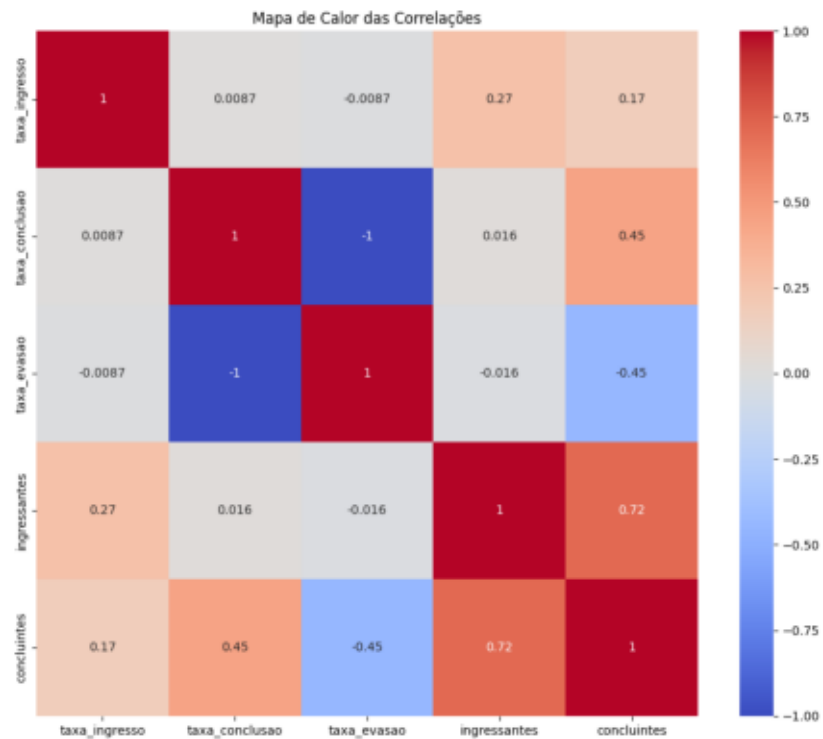




## Grafico Taxa Evasao Modalidade



## Grafico Mapa Calor Correlacoes



## **CONCLUSÃO GERAL**

- Taxa de evasão: Alta na maioria dos cursos, picos extremos em 100%
- Taxa de conclusão: Muito baixa em diversos cursos
- Relação ingressantes/concluintes: Correlação positiva, mas dispersa
- Modalidade de ensino: EAD com risco crítico de evasão
- Taxa de ingresso: Variação acentuada na ocupação de vagas
- Modelagem: Regressão linear binária com 100% acurácia — revisar