# **Análise dos Dados de Ingressantes e Formandos por Sexo (Diversos Centros)**

**Autores: Arthur Bogacki Veríssimo, Leandro Oliveira do Nascimento**

## **Introdução**

Nesta análise, investigamos os dados de ingressantes e formandos de diferentes cursos para verificar possíveis correlações entre esses dois fatores. Em particular, buscamos identificar se cursos com maior número de ingressantes apresentam também maior número de formandos, o que poderia sugerir um equilíbrio saudável. Por outro lado, se cursos com mais ingressantes não tiverem uma correspondência significativa em formandos, isso pode indicar uma alta taxa de evasão, possivelmente relacionada à dificuldade do curso.

Além disso, analisaremos os prédios com maior variação entre ingressantes e formandos, com o intuito de identificar áreas específicas mais afetadas pela desistência.

## **Pré-processamento dos Dados**

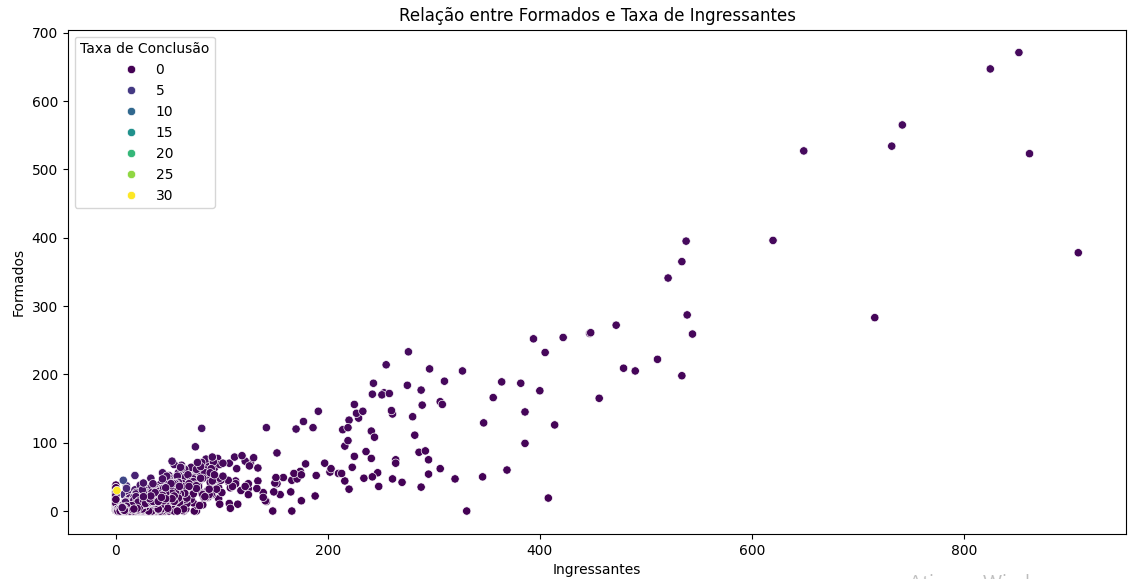
O primeiro passo do processo foi realizar o pré-processamento dos dados. Carregamos os dados de todas as tabelas, que inicialmente estavam separadas por prédio. Com isso, agrupamos todos os dados em uma única tabela, adicionando uma coluna chamada **“nome\_predio”** para identificar de qual prédio o curso pertencia. Também calculamos a **Taxa de Conclusão**, utilizando a fórmula:

Taxa de Conclusão = Formandos / Ingressantes

Além disso, removemos a coluna **“NIVEL\_CURSO”**, pois todos os cursos tinham o mesmo nível (graduação).

## **Gráfico: Relação entre Ingressantes e Formados**

A seguir, um gráfico que visualiza a relação entre ingressantes e formados, destacando os cursos que possuem uma maior quantidade de ingressantes e formandos:



## **Cálculo da Taxa de Conclusão e Taxa de Evasão**

Com os dados limpos e processados, o próximo passo foi calcular a **Taxa de Conclusão** e a **Taxa de Evasão**:

* **Taxa de Conclusão: Formandos\_Total / Ingressantes\_Total**

Taxa de Conclusão =

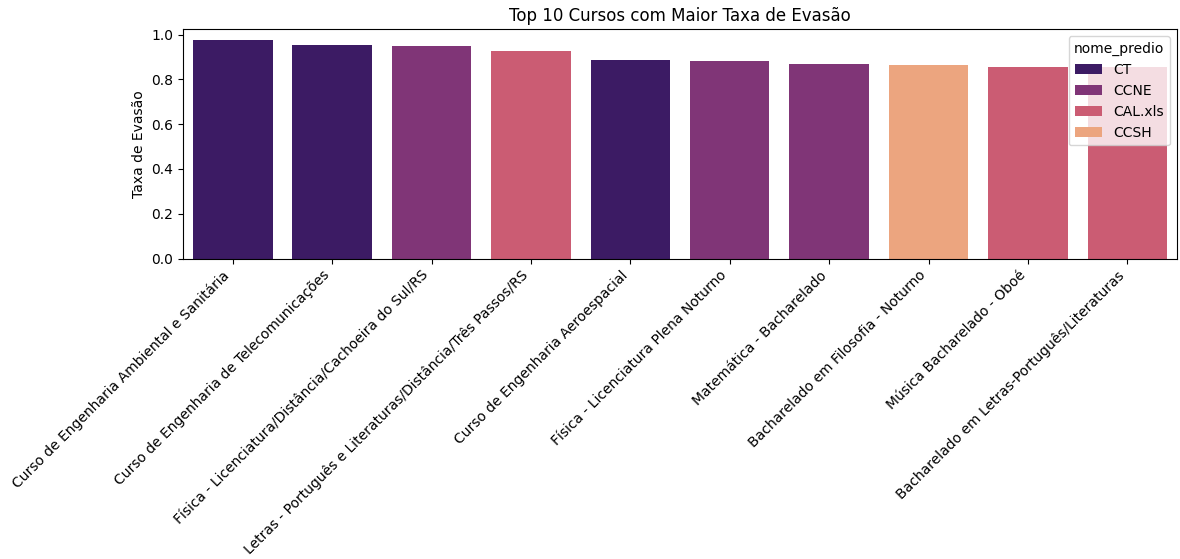
* **Taxa de Evasão: 1 - Taxa de Conclusão**

Além disso, optamos por **ignorar cursos com 0 ingressantes ou 0 formandos**. Isso foi feito para evitar distorções, já que esses dados podem representar erros de coleta ou cursos que já não existem mais.

## **Análise das Taxas de Evasão e Conclusão**

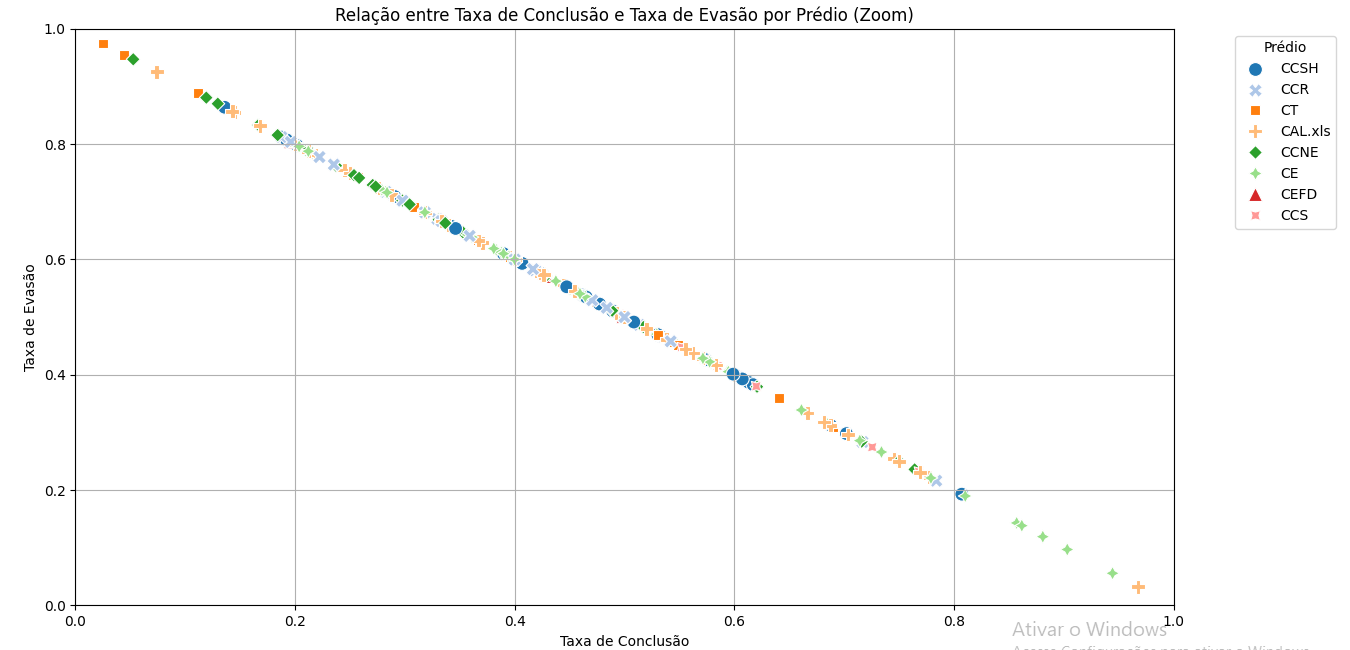
A partir dos cálculos, focamos na identificação dos **cursos com maior taxa de evasão**. Analisando os dados e os gráficos gerados, observamos que cursos de áreas como **exatas** (CT e CCNE) possuem uma maior taxa de desistência. Isso pode estar relacionado a diversos fatores, como a maior dificuldade das disciplinas dessas áreas.

Outro ponto a ser observado foi o **Centro de Artes e Letras (CAL)**, que também apresentou taxas elevadas de evasão, possivelmente devido à desvalorização das artes no Brasil, entre outros fatores.



## **Gráfico de Dispersão: Taxa de Conclusão vs. Taxa de Evasão**

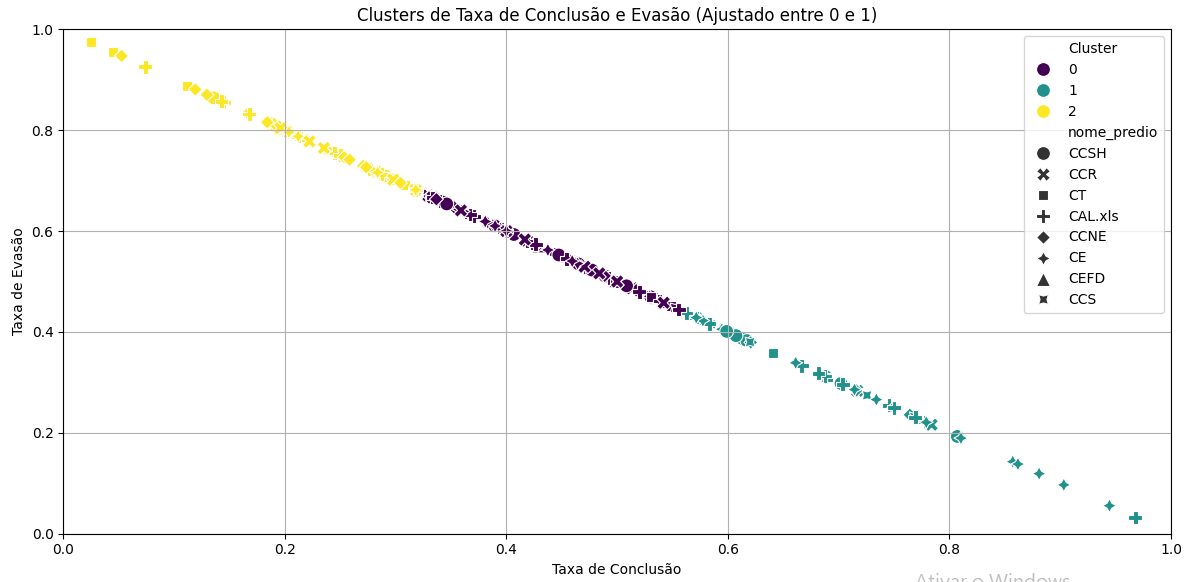
No gráfico abaixo, cada símbolo desenhado representa um curso, e as cores e formas geométricas indicam os prédios. Ao analisar a dispersão dos pontos, é possível observar que cursos com taxa de evasão próxima de 1 (mais ingressantes que formados) estão concentrados em prédios como o **CT** e o **CCNE**:



## **Agrupamento com Algoritmo K-Means**

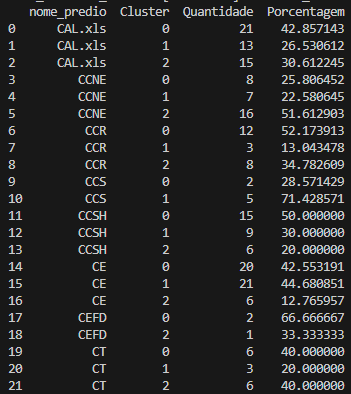
Utilizamos o algoritmo **K-Means** para agrupar os cursos em três grupos, considerando aqueles que tiveram taxa de conclusão entre 0 e 1. O agrupamento gerou os seguintes grupos:

* **Grupo 0 (ROXO):** Cursos com taxas equilibradas de ingresso e formação. Esses cursos têm números similares de ingressantes e formandos.
* **Grupo 1 (AZUL):** Cursos onde as taxas de conclusão são maiores que as de ingresso, ou seja, cursos com muitos formandos e poucos ingressantes.
* **Grupo 2 (AMARELO):** Cursos com alta taxa de evasão, onde o número de ingressantes é maior do que o número de formandos.



## **Tabela de Porcentagem por Prédio e Cluster**

A tabela a seguir mostra a porcentagem de cursos de cada prédio que pertence a cada grupo de cluster. Podemos observar que o **CT** tem uma distribuição equilibrada entre os três grupos, porém com 40% dos cursos obtendo maiores taxas de evasão, enquanto o **CCNE** tem uma alta concentração de cursos no **Grupo 2**, indicando uma taxa de evasão mais alta.



—----------------------------------------------------------------------------------------

## **Conclusões**

A análise dos dados revelou padrões interessantes sobre as taxas de evasão e conclusão nos cursos de diferentes centros. Os resultados indicam que:

1. **Áreas de Exatas (CT e CCNE):** Apresentam uma maior taxa de evasão, o que pode estar relacionado à dificuldade das disciplinas.
2. **Análise de Agrupamento:** O K-Means ajudou a identificar padrões em relação à taxa de evasão e conclusão, agrupando os cursos em três categorias com base nas suas características.

## **Considerações Finais**

Este estudo oferece uma visão importante sobre a relação entre ingressantes, formandos e evasão nos cursos, com implicações para o planejamento e políticas educacionais. A análise por clusters pode ajudar na identificação de cursos que exigem atenção especial em termos de taxa de evasão, e os resultados podem ser utilizados para estratégias de retenção de alunos.