

CENTRO UNIVERSITÁRIO IBMR

ÂNIMA EDUCAÇÃO

PEDRO AUGUSTO COSTA CORDEIRO CUNHA RIBEIRO

CONSTRUÇÃO DE UM *PIPELINE* DE DADOS PARA ANÁLISE DE *CHURN*

Rio de Janeiro

2024

PEDRO AUGUSTO COSTA CORDEIRO CUNHA RIBEIRO

CONSTRUÇÃO DE UM *PIPELINE* DE DADOS PARA ANÁLISE DE *CHURN*

Trabalho de Conclusão de Curso apresentado ao Curso de graduação de Engenharia de Produção em Engenharia de Produção, do Centro Universitário IBMR da Ânima educação, como requisito parcial para obtenção do título de Bacharel.

Orientador: Professor Michael Leone Madureira de Souza

Rio de Janeiro

2024

PEDRO AUGUSTO COSTA CORDEIRO CUNHA RIBEIRO

CONSTRUÇÃO DE UM *PIPELINE* DE DADOS PARA ANÁLISE DE *CHURN*

Este Trabalho de Conclusão de Curso foi julgado adequado à obtenção do título de Bacharel em Engenharia de Produção e aprovado em sua forma final pelo Curso de Engenharia de Produção, do Centro Universitário IBMR da Ânima Educação.

Rio de Janeiro, 03 de novembro de 2024.

Professor e orientador Michael Leone Madureira de Souza
Centro Universitário IBMR

Banca Orientadora
Centro Universitário IBMR

RESUMO

O presente trabalho tem como objetivo desenvolver e apresentar a construção de um pipeline de dados para a análise de churn em uma instituição de ensino superior, utilizando uma abordagem de análise por coortes. A proposta envolve a criação de uma base de dados simulada de 10 anos, abrangendo o período de 2013 a 2023, com informações relevantes como matrículas, cancelamentos, cursos, e motivos de evasão. O pipeline é responsável por realizar o tratamento, a limpeza e a integração dos dados, permitindo uma análise descritiva e preditiva da retenção de alunos. Os resultados obtidos demonstram a importância de um pipeline de dados bem estruturado para facilitar a tomada de decisões estratégicas voltadas à retenção de estudantes e melhoria contínua de processos institucionais.

Palavras-chave: *Churn. Pipeline. Análise de Cohort. Análise de dados. Engenharia de dados*

ABSTRACT

This work aims to develop and present the construction of a data pipeline for churn analysis in a higher education institution, using a cohort analysis approach. The proposal involves creating a simulated database of 10 years, covering the period from 2013 to 2023, with relevant information such as enrollments, cancellations, courses, and dropout reasons. The pipeline is responsible for data processing, cleaning, and integration, allowing descriptive and predictive analysis of student retention. The results demonstrate the importance of a well-structured data pipeline to facilitate strategic decision-making aimed at student retention and continuous improvement of institutional processes. Keywords: churn, data pipeline, cohort analysis, student retention, higher education.

Keywords: Churn. Pipeline. Cohort Analysis. Data analytics. Data engineering

SUMÁRIO

1 INTRODUÇÃO	8
1.1 OBJETIVO	9
1.2 METODOLOGIA	10
2 REVISÃO BIBLIOGRÁFICA	11
3 FUNDAMENTAÇÃO TEÓRICA	13
3.1 <i>PIPELINE</i> DE DADOS	13
3.2 <i>CHURN</i>	13
3.3 CÁLCULO DE <i>CHURN</i>	14
3.4 ANÁLISE DE <i>COHORT</i>	14
4 APRESENTAÇÃO DE RESULTADOS	16
4.1 <i>PIPELINE</i> DE DADOS EM PYTHON	16
4.2 DASHBOARD EM POWER BI	16
4.3 MATRIZ DE COHORT EM EXCEL	22
4.4 TAXAS DE CHURN	23
5 CONSIDERAÇÕES FINAIS	25
REFERÊNCIAS	26

6 ANEXOS	28
6.1 REPOSITÓRIO NO GITHUB	28

1 INTRODUÇÃO

Na contemporaneidade empresarial, onde a informação se torna vital para o êxito organizacional, a Engenharia de Dados surge como um pilar essencial para a sustentabilidade operacional. A integração estratégica de ferramentas de gestão de dados tem sido uma abordagem crucial para otimizar processos, além de proporcionar uma visão mais aprimorada e abrangente das operações.

Nos últimos anos, a análise de dados evoluiu de uma simples prática operacional para se consolidar como uma ferramenta estratégica indispensável para as organizações. Com o crescimento exponencial de dados gerados diariamente, a capacidade de coletar, processar e interpretar informações de forma eficiente tem proporcionado uma vantagem competitiva significativa. A análise de dados permite insights profundos sobre o comportamento dos consumidores, otimiza processos internos, melhora a eficiência operacional e apoia a inovação. Além disso, ferramentas avançadas de análise preditiva possibilitam a antecipação de tendências e a identificação de oportunidades e riscos com maior precisão. Em um ambiente de negócios cada vez mais dinâmico e competitivo, a habilidade de transformar dados em decisões assertivas tornou-se um diferencial crucial para o sucesso e a sustentabilidade das organizações a longo prazo.

No contexto educacional, o gerenciamento eficaz de alunos e a fidelização têm ganhado destaque, considerando o aumento da concorrência entre instituições e a necessidade de otimização dos recursos acadêmicos. Com a crescente digitalização do setor, as instituições de ensino têm acesso a um volume significativo de dados que, se bem tratados e analisados, podem fornecer insights valiosos sobre o comportamento e as necessidades dos alunos.

A taxa de *churn* impacta diretamente a rentabilidade das instituições de ensino, sendo um indicador crucial da satisfação dos alunos e da eficiência das estratégias de marketing e retenção. No entanto, a ausência de um pipeline de dados eficiente pode dificultar a identificação de padrões e a compreensão das razões que levam os alunos a abandonarem o curso.

Diante desse cenário, o desenvolvimento de um pipeline de dados robusto torna-se essencial para monitorar e analisar métricas de evasão, permitindo que as instituições identifiquem as turmas mais propensas ao cancelamento e adotem estratégias proativas de retenção. A implementação de processos eficazes de coleta, tratamento e análise de dados possibilita a segmentação dos alunos e a personalização de ofertas, aumentando as chances de retenção.

Este trabalho tem como objetivo explorar a construção de um pipeline de dados simulados para análise de *churn* utilizando a metodologia de *cohort*, com foco em transformar dados brutos em informações estratégicas. Serão abordados os benefícios da análise descritiva e a importância de um processo estruturado na tomada de decisões, com o intuito de otimizar as estratégias de retenção e melhorar a experiência do aluno no setor educacional.

1.1 OBJETIVO

O objetivo deste trabalho é desenvolver um pipeline de dados eficiente para análise de *churn* em uma instituição de ensino superior, utilizando uma base de dados simulada para identificar padrões de cancelamento e auxiliar na tomada de decisões estratégicas de retenção de alunos.

Os objetivos específicos são:

- a)** Criar e estruturar uma base de dados simulada de matrículas e cancelamentos de alunos ao longo de 10 anos;
- b)** Construir um pipeline de dados;
- c)** Aplicar técnicas de análise de *cohort* para avaliar a retenção e cancelamento de alunos;
- d)** Gerar visualizações que facilitem a interpretação dos resultados e a implementação de estratégias de retenção;
- e)** Validar a eficácia do pipeline e dos métodos utilizados para análise de *churn*.

1.2 METODOLOGIA

O presente trabalho foi conduzido com uma abordagem prática, centrada na criação de um pipeline de dados para análise de *churn* no contexto educacional.

Inicialmente, foi desenvolvido um código em Python, utilizando as bibliotecas *os*, *pandas*, *numpy*, *random* e *datetime*, com ênfase especial em *pandas* e *numpy* devido à sua eficiência no tratamento e manipulação de dados. A partir desse código, foram simulados dados fictícios referentes a um período de 10 anos, gerando mensalmente relatórios em formato CSV. Esses relatórios continham informações detalhadas sobre matrícula, como curso, grupo, status, cancelamento, gênero, idade, motivo de cancelamento, modalidade, turno, mensalidade, desconto, valor de pagamento, data de matrícula e data de cancelamento.

Na sequência, estruturou-se um pipeline de dados que permitiu a concatenação de múltiplos relatórios avulsos em um único arquivo consolidado. Este arquivo foi preparado para integração com o Power BI, onde foi desenvolvido um dashboard com o objetivo de facilitar a identificação de perfis de alunos com maior propensão ao cancelamento. A consolidação dos dados eliminou a necessidade de etapas adicionais de tratamento, uma vez que os dados foram gerados de maneira estruturada e ajustada às necessidades da análise.

O dashboard no Power BI foi elaborado utilizando gráficos de colunas, barras e rosca, permitindo a visualização clara de padrões de cancelamento. Os dados simulados, sendo controlados, evidenciaram tendências de forma explícita, o que contribuiu para uma análise direcionada. Adicionalmente, o Power BI foi utilizado para criar agrupamentos e visualizações adicionais, resultando em tabelas que destacavam as aberturas mais relevantes e a quantidade de alunos distintos agrupados por essas categorias.

Por fim, uma matriz de *cohort* foi desenvolvida no Excel, aplicando-se fórmulas e atalhos que facilitavam a filtragem e análise de retenção e cancelamento ao longo do tempo. A matriz foi configurada para permitir a atualização automática das análises à medida que novos dados fossem inseridos, garantindo a eficiência na revisão contínua das informações.

2 REVISÃO BIBLIOGRÁFICA

A análise de *churn*, amplamente estudada no setor empresarial, refere-se à taxa de abandono de clientes ou usuários de um serviço. Segundo Fader (2012), entender o *churn* é essencial para desenvolver estratégias de retenção eficazes, uma vez que a perda de clientes impacta diretamente a receita e a sustentabilidade das organizações. No contexto educacional, o *churn* se manifesta como evasão escolar, onde estudantes deixam de se matricular ou abandonam seus cursos, prejudicando tanto sua trajetória acadêmica quanto o desempenho institucional (Silva et al., 2020).

Vários estudos têm explorado a aplicação de técnicas analíticas para prever e mitigar o *churn*. No setor de saúde, Martins e Silva (2020) utilizaram um modelo preditivo para identificar clientes de planos de saúde com maior risco de cancelamento. Esse estudo demonstrou como a análise de *churn* pode ser uma ferramenta poderosa para a retenção, ao permitir intervenções direcionadas com base em dados históricos e comportamentais (Martins & Silva, 2020).

No campo do marketing, a análise de *cohort* tem sido amplamente utilizada para segmentar e entender o comportamento de grupos específicos de clientes ao longo do tempo. Fader (2012) explica que essa técnica permite que as organizações rastreiem como diferentes coortes reagem a variáveis específicas, como mudanças em produtos ou campanhas promocionais, oferecendo insights valiosos para a tomada de decisões estratégicas.

A aplicação dessas metodologias no setor educacional tem sido explorada por instituições como o Instituto Federal do Rio Grande do Norte (IFRN) e o Instituto Federal da Paraíba (IFPB). No IFRN, Barros et al. (2020) implementaram um pipeline de dados para prever a evasão escolar, utilizando técnicas de modelagem preditiva para identificar estudantes em risco. A análise de *cohort* foi empregada para segmentar os estudantes com base em diferentes variáveis, como ano de ingresso e curso, permitindo uma análise detalhada das taxas de retenção e evasão ao longo do tempo (Barros et al., 2020).

Silva et al. (2020) no IFPB adotaram uma abordagem semelhante, utilizando um pipeline de dados para analisar os fatores associados à evasão escolar. A análise de *cohort* revelou padrões de comportamento em diferentes coortes de estudantes,

destacando a importância de variáveis socioeconômicas e acadêmicas na decisão de permanecer ou abandonar o curso. Os insights obtidos permitiram a implementação de estratégias de retenção mais eficazes, como programas de apoio financeiro e acadêmico (Silva et al., 2020).

Esses estudos demonstram que, assim como no setor empresarial, a análise de *churn* e *cohort* no contexto educacional oferece uma abordagem robusta para entender e melhorar a retenção de alunos. A integração de pipelines de dados, bom gerenciamento de banco de dados, junto com técnicas analíticas não apenas facilita a identificação de padrões de evasão, mas também permite o desenvolvimento de intervenções proativas e personalizadas, contribuindo para a melhoria contínua das práticas institucionais.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 PIPELINES DE DADOS

Um pipeline de dados consiste em uma série de processos automatizados que viabilizam a coleta, o tratamento e a análise de dados de maneira contínua e eficiente. A estruturação adequada de um pipeline é essencial para assegurar a qualidade e a relevância dos dados processados, permitindo análises mais precisas e informativas.

A construção de um pipeline robusto geralmente envolve três etapas principais: extração (*extract*), transformação (*transform*) e carregamento (*load*) (ETL). Durante a extração, os dados são coletados de diversas fontes, como bancos de dados, APIs e arquivos. A fase de transformação é dedicada à limpeza, normalização e preparação dos dados para análise. Finalmente, os dados transformados são carregados em sistemas de armazenamento, como *Data Warehouses*, onde podem ser acessados para análises futuras.

Zikopoulos et al. (2012) destacam que pipelines de dados eficazes são cruciais para iniciativas de *business intelligence*, pois permitem a captura e tratamento de dados, facilitando análises contínuas e a geração de insights acionáveis.

No contexto deste trabalho, o pipeline foi utilizado para gerar, processar e compactar dados simulados de matrículas e cancelamentos de alunos, assegurando a integridade dos dados e sua integração eficiente com ferramentas de visualização e análise.

3.2 CHURN

O conceito de *churn*, ou taxa de evasão, refere-se ao abandono ou cancelamento de um serviço por parte dos clientes. Esse indicador é amplamente utilizado para medir a lealdade e a retenção dentro de organizações. Kumar e Petersen (2012) destacam a importância de modelos preditivos na identificação precoce de padrões de *churn*, permitindo que as organizações implementem estratégias preventivas de retenção.

No contexto educacional, o *churn* é evidenciado pela desistência de alunos, seja por não renovarem suas matrículas ou por abandonarem o curso. A compreensão das causas do *churn* educacional é crucial, dado que a perda de alunos não apenas

afeta financeiramente as instituições, mas também impacta sua reputação e capacidade de atrair novos estudantes. Fatores como insatisfação com o ensino, falta de engajamento e mudanças nas circunstâncias pessoais são frequentemente citados como razões para a evasão, ressaltando a necessidade de intervenções direcionadas para mitigar esses fatores.

3.3 CÁLCULO DE *CHURN*

A taxa de *churn* é uma métrica crucial para a avaliação da retenção, seja em empresas ou instituições educacionais. O *churn* pode ser calculado usando a seguinte equação:

Figura 1 – Equação de *churn*

$$Churn = \left(\frac{Cc}{Ct} \right) \times 100$$

Cc = Número de clientes que cancelaram durante o período.

Ct = Número total de clientes no início do período.

Este cálculo fornece uma visão clara sobre a saúde da base de clientes ou alunos de uma instituição, permitindo a implementação de estratégias direcionadas para minimizar o *churn*. No ambiente educacional, a aplicação dessa métrica é fundamental para monitorar a eficácia de iniciativas de retenção e para ajustar estratégias conforme necessário.

3.4 ANÁLISE DE *COHORT*

A análise de *cohort* é uma técnica estatística utilizada para examinar o comportamento de grupos específicos de indivíduos, denominados coortes, que compartilham características ou experiências comuns ao longo do tempo. Fader (2012) destaca que essa abordagem permite identificar tendências e padrões que podem não ser evidentes em análises globais.

A aplicação dessa técnica envolve várias etapas. Primeiro, a definição do coorte, que pode ser baseada em critérios como a data de ingresso ou características demográficas. Em seguida, os dados relevantes são coletados para cada *cohort*. A análise temporal é então conduzida para avaliar como o comportamento de cada grupo evolui ao longo do tempo, revelando padrões de retenção e *churn*.

O auxílio na segmentação eficaz da base de clientes ou usuários, proporciona insights que informam decisões estratégicas e a personalização de ofertas. No contexto educacional, essa técnica é particularmente útil para entender a dinâmica de retenção de estudantes e para identificar oportunidades de melhoria nos processos institucionais.

Neste trabalho, a análise de *cohort* foi utilizada para segmentar os alunos simulados e avaliar suas taxas de retenção e cancelamento ao longo de safras.

4 APRESENTAÇÃO DE RESULTADOS

A análise de dados realizada neste trabalho revelou insights cruciais sobre o comportamento dos alunos e suas taxas de retenção. Esta seção apresenta os principais resultados obtidos através do pipeline de dados implementado em Python, das visualizações geradas no Power BI e da matriz de *cohort* feita em Excel. Todos esses recursos estão disponíveis em um repositório no GitHub, acessível no seguinte endereço: <https://github.com/PAugustoRibeiro/TCC>.

4.1 PIPELINE DE DADOS EM PYTHON

O pipeline de dados desenvolvido em Python mostrou-se altamente eficaz na geração, consolidação e preparação dos dados simulados para análise. Utilizando as bibliotecas *random* e *datetime*, foi possível criar dados aleatórios, porém coerentes, aplicando regras de probabilidade e restrições temporais para garantir a consistência das datas.

Com o suporte de *pandas* e *numpy*, grandes volumes de dados foram processados de maneira eficiente, permitindo a manipulação e estruturação necessárias para a análise. Essas bibliotecas facilitaram o tratamento dos dados, assegurando que todas as informações fossem organizadas de forma consistente e pronta para uso.

Além disso, a biblioteca os desempenhou um papel crucial no gerenciamento do sistema de arquivos. Essa etapa permitiu consolidar múltiplos arquivos CSV em uma pasta unificada, garantindo um fluxo de dados organizado e centralizado.

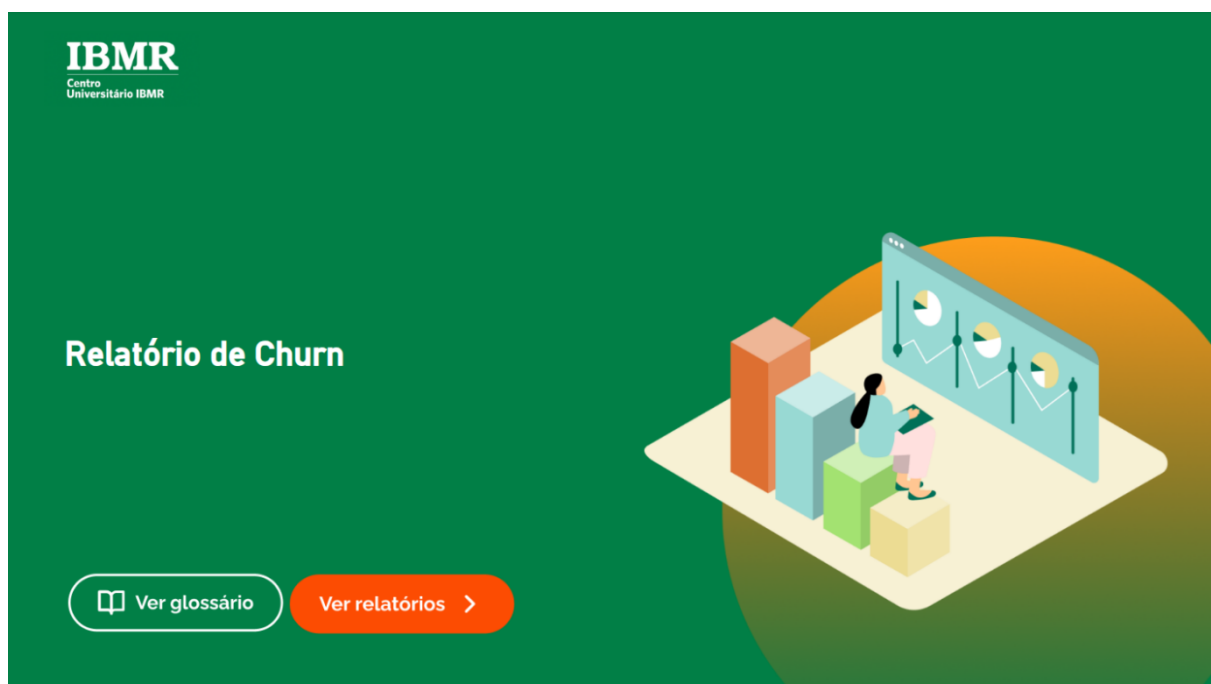
Por fim, os dados foram integrados em um único arquivo CSV, o que simplificou sua importação e análise no Power BI. Esse processo assegurou a integridade e a consistência dos dados, proporcionando uma base sólida para análises precisas e confiáveis.

4.2 DASHBOARD EM POWER BI

O dashboard desenvolvido no Power BI forneceu uma visão abrangente das taxas de retenção e cancelamento, categorizadas por grupo, modalidade, motivo de cancelamento e turno. A interface interativa permitiu explorar diversos insights,

facilitando a identificação de padrões e tendências que influenciam o comportamento dos alunos.

Figura 2 – Capa do relatório



A página inicial do relatório apresenta uma visão geral intuitiva, destacando o tema central do relatório. Esta página contém dois botões interativos: o botão "Ver glossário", que direciona ao repositório de arquivos no GitHub: <https://github.com/PAugustoRibeiro/TCC>, e o botão "Ver relatórios", que proporciona acesso direto ao *dashboard* de análise. Essa estrutura facilita a navegação e oferece ao usuário um ponto de partida claro para explorar os dados e os insights disponíveis.

Figura 3 - Dashboard



A análise detalhada apresentada no dashboard interativo fornece insights fundamentais sobre as taxas de cancelamento e suas diversas dimensões. Sendo os principais achados:

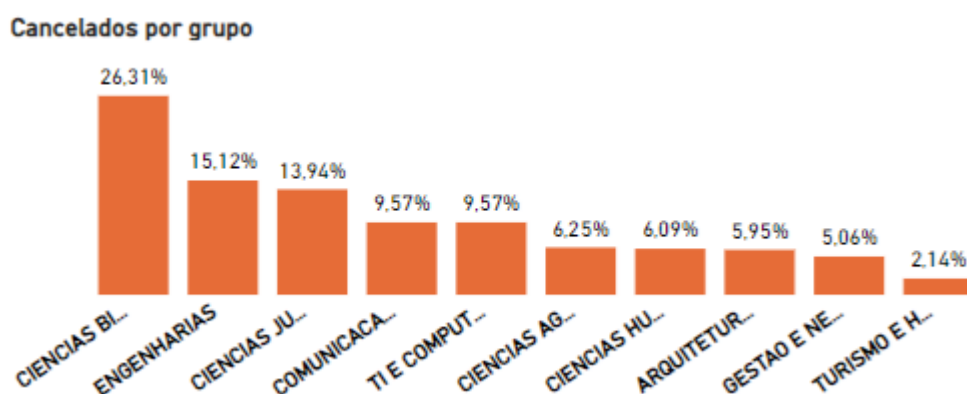
- Curso: Os gráficos de colunas destacam as taxas de cancelamento distribuídas por curso. Notavelmente, não se observa uma tendência clara que favoreça ou prejudique cursos específicos, indicando que o cancelamento está uniformemente distribuído entre os diferentes cursos oferecidos.

Figura 4 – Gráfico de barras de cancelados por curso



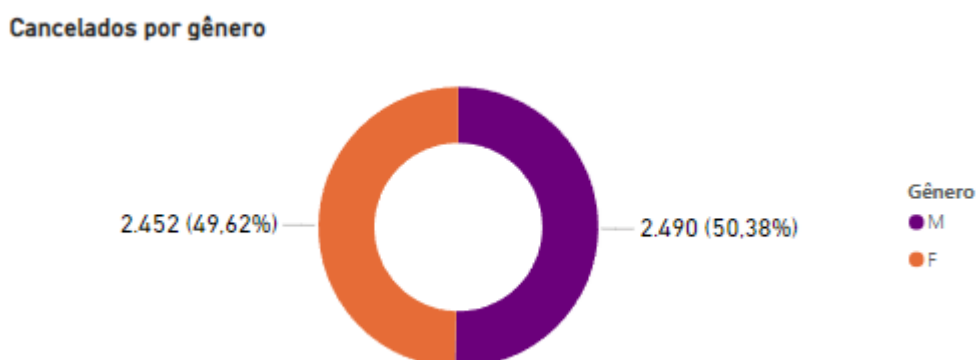
- Grupo: A análise por grupos revela áreas com maiores taxas de cancelamento. Grupos como Ciências Biológicas e da Saúde apresentam a maior taxa de cancelamento, com 26,31%, seguidos por Engenharias com 15,12%, e Ciências Jurídicas com 13,94%. Esses resultados indicam que essas áreas requerem maior atenção para estratégias de retenção.

Figura 5 – Gráfico de colunas de cancelados por grupo



- Gênero: O gráfico de rosca mostra uma distribuição equilibrada de cancelamentos entre os gêneros, com 49,62% no gênero feminino e 50,38% no masculino. Essa paridade indica que o gênero não é um fator determinante nas taxas de cancelamento.

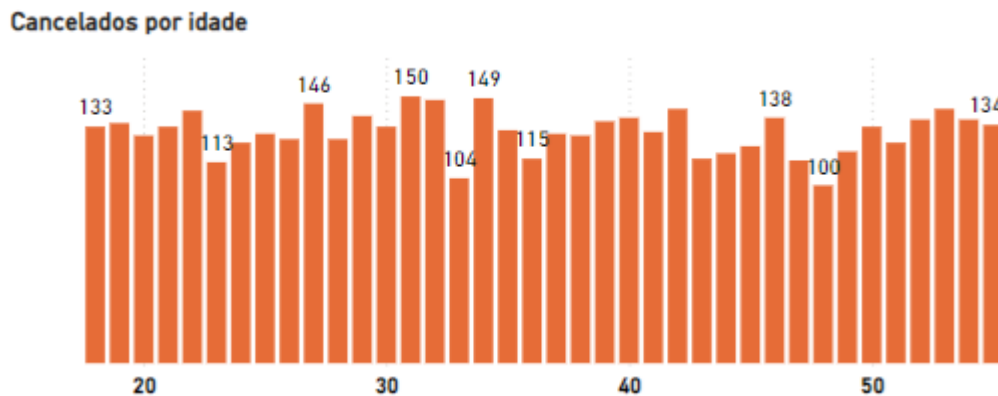
Figura 6 – Gráfico de rosca de cancelados por gênero



- Idade: A análise por faixa etária revela que as taxas de cancelamento são igualmente distribuídas entre diferentes grupos de idade, sem

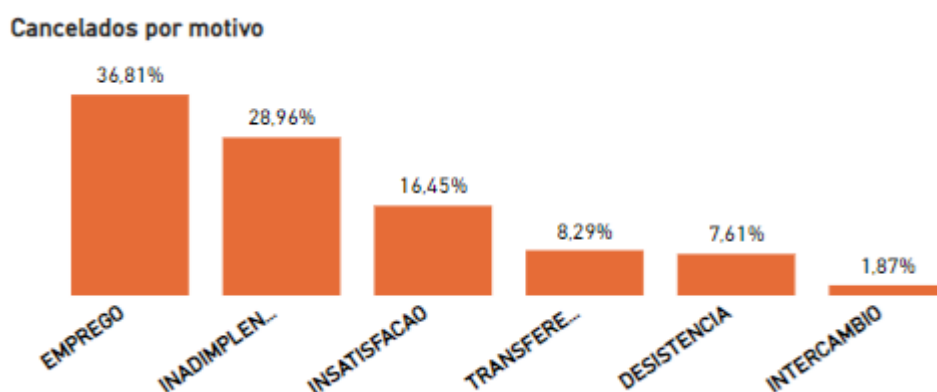
tendências significativas. Isso sugere que a idade não é um fator influente no comportamento de cancelamento.

Figura 7 – Gráfico de colunas de cancelados por idade



- **Motivos de Cancelamento:** Os principais motivos de cancelamento foram classificados em um gráfico de barras. Os dados mostram que emprego é o principal motivo, responsável por 36,81% dos cancelamentos, seguido por inadimplência com 28,96%, e insatisfação com 16,45%. Esses insights são cruciais para a instituição abordar problemas específicos que levam ao desligamento dos alunos.

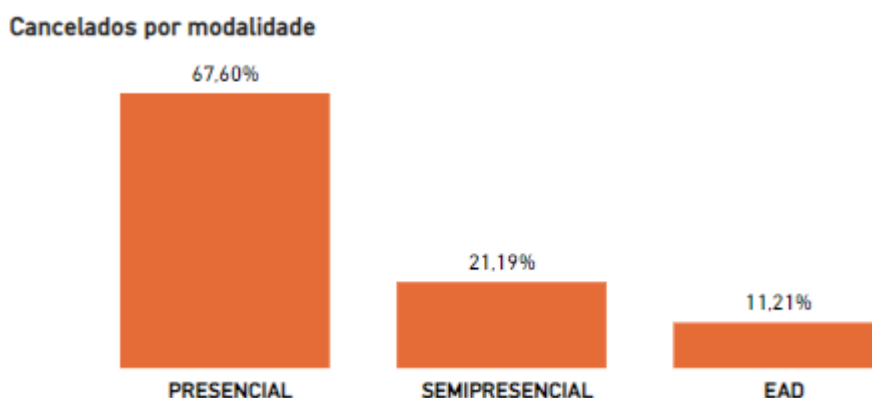
Figura 8 – Gráfico de colunas de cancelados por motivo de cancelamento



- **Modalidade:** A modalidade de ensino também influencia nas taxas de cancelamento. O maior índice está na modalidade presencial com 67,6%, seguida pela modalidade semipresencial com 21,19%, e EAD

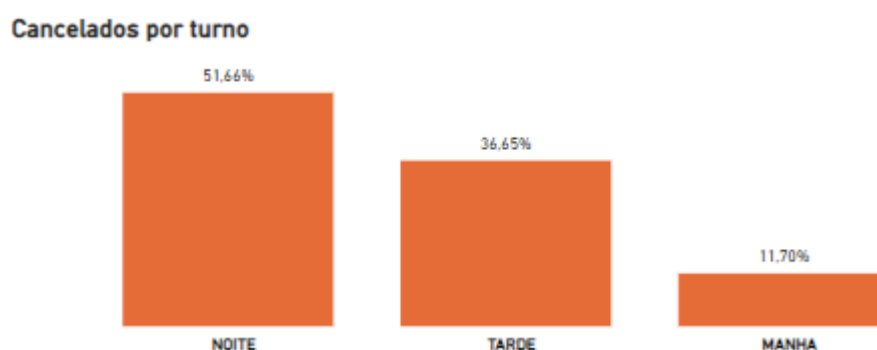
com 11,21%. Esses dados ajudam a identificar onde a experiência do aluno pode ser melhorada.

Figura 9 – Gráfico de colunas de cancelados por modalidade



- Turno: Os dados indicam que o turno noite apresenta a maior taxa de cancelamento com 51,66%, seguido pela tarde com 36,65%, e manhã com 11,7%. Esses resultados sugerem que os alunos do turno da noite podem estar enfrentando mais desafios que contribuem para o cancelamento.

Figura 10 – Gráfico de colunas de cancelados por turno

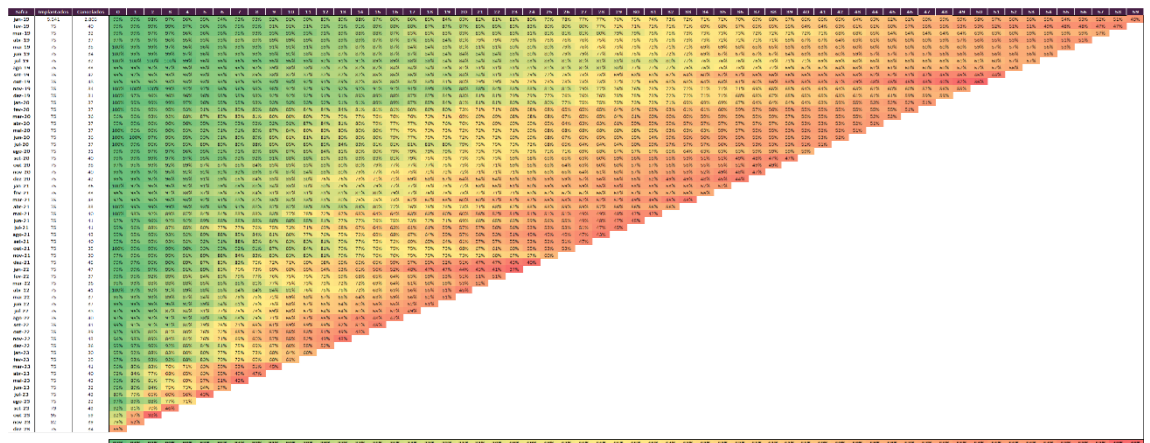


Este dashboard foi essencial para identificar padrões e tendências em cancelamentos. Ele permitiu a definição de quais aberturas são mais relevantes para a análise de *cohort*, ajudando na compreensão de fatores críticos que afetam a retenção de alunos. A aplicação dessas análises possibilita o desenvolvimento de estratégias mais eficazes para a retenção, melhorando tanto a experiência do aluno quanto os resultados institucionais.

4.3 MATRIZ DE COHORT EM EXCEL

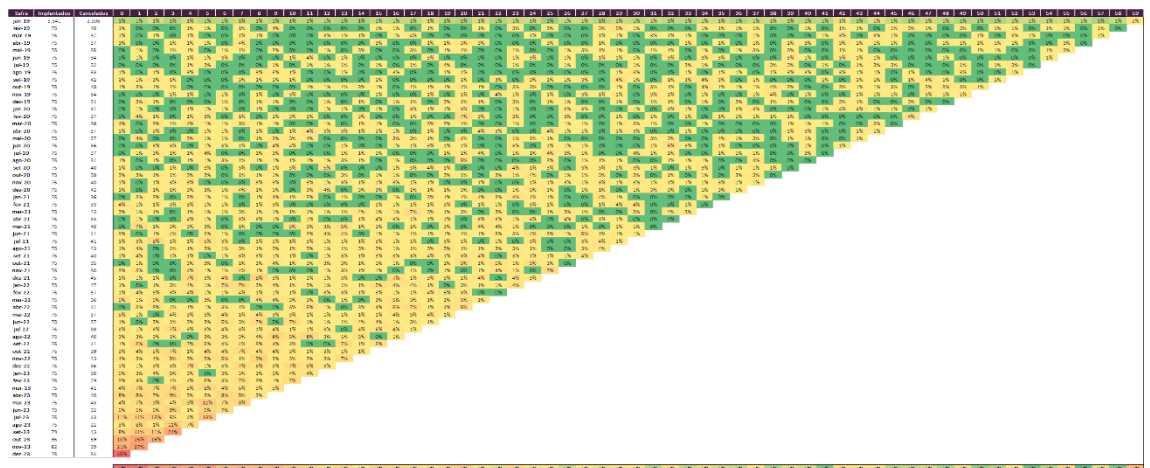
A matriz de cohort, elaborada no Excel, desempenhou um papel crucial na análise de padrões de retenção e cancelamento ao longo dos anos. Essa matriz segmenta os alunos em coortes baseadas no ano e mês de ingresso, permitindo uma análise detalhada das taxas de retenção e cancelamento mês a mês.

Figura 11 – Matriz de *cohort* de retenção



A matriz de retenção, como ilustrado acima, evidencia o percentual de alunos que permanecem matriculados ao longo do tempo. As células coloridas variam do verde ao vermelho, indicando, respectivamente, altas e baixas taxas de retenção. Esta visualização facilita a identificação de tendências ao longo do tempo, destacando coortes com melhor desempenho.

Figura 12 – Matriz de *cohort* de churn



A matriz de *churn* complementa a análise ao mostrar a proporção de alunos que cancelaram suas matrículas. A coloração semelhante, porém inversa do verde ao vermelho, ajuda a identificar rapidamente os períodos críticos em que as taxas de cancelamento são mais elevadas.

A análise revelou tendências importantes, como:

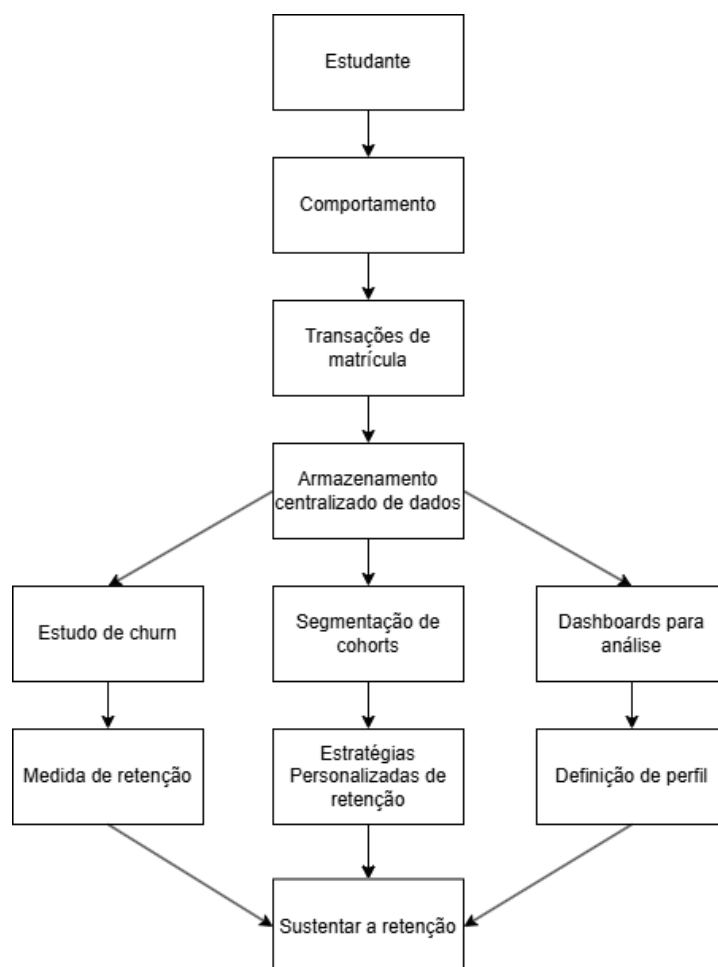
- Padrões de cancelamento: Alunos que ingressaram em anos mais recentes demonstraram taxas de retenção inferiores, sugerindo melhorias nos programas de apoio e adaptação das estratégias institucionais.

4.4 TAXAS DE *CHURN*

A taxa de churn calculada para o período analisado foi de 49,92%, o que representa 4.942 alunos que cancelaram suas matrículas dentro de uma carteira total de 9.900 alunos. Essa métrica reflete a proporção de alunos que optaram por cancelar em relação ao total de matrículas no início do período. A análise das causas desse abandono foi essencial para compreender as dinâmicas de retenção e para subsidiar a formulação de estratégias de retenção mais eficazes.

O fluxo de processos apresentado a seguir ilustra as etapas envolvidas na análise e cálculo da taxa de churn, desde a coleta e estruturação dos dados até a visualização e interpretação dos resultados.

Figura 13 – Fluxo de processos



5 CONSIDERAÇÕES FINAIS

Este trabalho apresentou a importância do desenvolvimento de um pipeline de dados para a análise de *churn* em instituições de ensino. Através da implementação de ferramentas de visualizações de dados e técnicas de análise de *cohort*, foi possível identificar padrões de perfil e comportamento que impactam diretamente o cancelamento dos alunos.

Os resultados obtidos destacam a necessidade de um suporte acadêmico mais robusto e a implementação de estratégias específicas para diferentes grupos, modalidades e turnos. A utilização de ferramentas de visualização de dados também se mostrou eficaz para a comunicação das métricas de retenção à equipe acadêmica.

Para trabalhos futuros, recomenda-se a continuidade da análise de dados e a inclusão de novas variáveis, como feedback dos alunos e taxas de participação em atividades extracurriculares. Essas informações podem contribuir ainda mais para a formulação de estratégias de retenção mais eficazes, melhorando a experiência acadêmica e a sustentabilidade financeira das instituições. Além disso, o uso de técnicas de aprendizado de máquina pode ser explorado para prever o *churn* com maior precisão, permitindo intervenções mais rápidas e personalizadas.

Em suma, a análise de *churn* é fundamental para que as instituições de ensino compreendam melhor o comportamento dos alunos, aumentem a taxa de retenção e, conseqüentemente, fortaleçam sua posição no mercado educacional. O sucesso da retenção de alunos não depende apenas de intervenções pontuais, mas de um compromisso contínuo com a qualidade da educação e com o apoio ao aluno ao longo de sua trajetória acadêmica.

REFERÊNCIAS

- 1 BARROS, J.; SILVA, F.; OLIVEIRA, R. Processo orientado a dados para a geração de modelos preditivos de evasão escolar. Instituto Federal do Rio Grande do Norte, 2020. Disponível em: https://repositorio.ufrn.br/jspui/bitstream/123456789/31933/1/Processoorientadodados_Barros_2020.pdf. Acesso em: 3 nov. 2024.
- 2 FADER, P. S. Customer Centricity: Focus on the Right Customers for Strategic Advantage. Wharton Digital Press, 2012. Disponível em: https://www.google.com.br/books/edition/Customer_Centricity/mwI2EAAAQBAJ?hl=pt-BR&gbpv=1&dq=FADER,+P.+S.+Customer+Centricity:+Focus+on+the+Right+Customers+for+Strategic+Advantage&printsec=frontcover. Acesso em: 3 nov. 2024.
- 3 IMDS Brasil. Diagnóstico de abandono e evasão escolar no Brasil. Instituto Mobilidade e Desenvolvimento Social, 2022. Disponível em: <https://imdsbrasil.org/doc/ImdsA001-2022-Diagn%C3%B3sticoAbandonoEvas%C3%A3oEscolarBrasil.pdf>. Acesso em: 3 nov. 2024.
- 4 IMDS Brasil. Evasão escolar segundo a PNADC 2022. Instituto Mobilidade e Desenvolvimento Social, 2023. Disponível em: <https://imdsbrasil.org/doc/ImdsNT001-2023-Evas%C3%A3oEscolarPNADC2022.pdf>. Acesso em: 3 nov. 2024.
- 5 MARTINS, A.; SILVA, M. Aplicação de Modelos Preditivos na Análise de Churn no Setor de Saúde. Revista de Saúde e Tecnologia, 2020. Disponível em: <https://repositorio.ifpb.edu.br/handle/177683/3206>. Acesso em: 3 nov. 2024.
- 6 PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO (PNUD). Evasão escolar no ensino médio atinge meio milhão de jovens por ano e perpetua desigualdade, alerta estudo do PNUD e FIRJAN Sesi. 2023. Disponível em: <https://www.undp.org/pt/brazil/news/evasao-escolar-no-ensino-medio-atinge-meio-milhao-de-jovens-por-ano-e-perpetua-desigualdade-alerta-estudo-do-pnud-e-firjan-sesi>. Acesso em: 3 nov. 2024.

7 SILVA, F.; BARROS, J.; OLIVEIRA, R. Análise de variáveis associadas à evasão escolar no IFPB. Instituto Federal da Paraíba, 2020. Disponível em: <https://repositorio.ifpb.edu.br/handle/177683/3188>. Acesso em: 3 nov. 2024.

8 Universidade Federal do Rio Grande do Sul (UFRGS). Análise da evasão escolar no Brasil. Revista Informação e Educação: Teoria e Prática, 2020. Disponível em: <https://seer.ufrgs.br/index.php/InfEducTeoriaPratica/article/view/87880>. Acesso em: 3 nov. 2024.

9 GLOBO. IBGE mede o problema nacional da evasão escolar. Jornal Nacional, 2020. Disponível em: <https://g1.globo.com/jornal-nacional/noticia/2020/07/15/ibge-mede-o-problema-nacional-da-evasao-escolar.ghtml>. Acesso em: 3 nov. 2024.

6 ANEXOS

6.1 REPOSITÓRIO NO GITHUB

Os recursos adicionais utilizados neste trabalho, incluindo o dashboard em Power BI, a matriz de *cohort* em Excel e o código em Python, estão disponíveis no seguinte repositório:

<https://github.com/PAugustoRibeiro/TCC>.