

FUNDAÇÃO ESCOLA DE COMÉRCIO ÁLVARES PENTEADO  
FECAP

ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

João Pedro Brosselin de Albuquerque Souza - 24026155

Marcella Santana Gonçalves Diniz Rocha - 24025750

Thays Helyda da Silva Pontes - 24026610

Gustavo de souza Castro - 20021558

Exploração de Dados

Fidelize

São Paulo - SP

2025

# Introdução

Nesta primeira etapa do projeto, foi realizada uma Análise Exploratória de Dados (AED) sobre um conjunto de dados simulados da plataforma Cannoli. O objetivo foi entender a estrutura, a qualidade e os padrões básicos dos dados para embasar a construção de um futuro dashboard interativo.

# 1. Etapas realizadas até o momento

## Carga e Inspeção Inicial

Os quatro conjuntos de dados (`Customers`, `Campaigns`, `Orders`, `CampaignQueue`) foram carregados a partir de arquivos CSV. Para cada um, foi realizada uma inspeção inicial para verificar a estrutura, os tipos de dados, a presença de valores nulos e a existência de registros duplicados.

## Limpeza e Preparação de Dados

Foi identificada a necessidade de um tratamento mais robusto nos dados de pedidos. As seguintes ações foram tomadas:

- **Valores Monetários:** A coluna `totalAmount` foi corretamente convertida de um formato textual (com `.` para milhar e `,` para decimal) para um formato numérico padrão, garantindo a precisão dos cálculos financeiros.
- **Datas:** A coluna `createdAt` foi convertida para um formato de data (`datetime`), especificando o formato brasileiro (`dayfirst=True`) para evitar ambiguidades.
- **Análise Exploratória e Visualização:** Com os dados limpos, diversas análises foram conduzidas e visualizadas através de gráficos, incluindo:
  - Análise da distribuição de gênero dos clientes.
  - Análise temporal (mensal e diária) do volume de pedidos e do faturamento.
  - Criação de gráficos avançados com eixo duplo para comparar diretamente as tendências de volume e faturamento.

# 2. Insights e Interpretações até o momento

A análise dos dados permitiu a extração de conclusões relevantes sobre o desempenho operacional e o comportamento do consumidor, que são cruciais para a tomada de decisões estratégicas. Os principais insights são:

## Sazonalidade do Negócio

A análise temporal mensal revela um claro padrão de sazonalidade nas operações. Foram identificados meses de alta e baixa tanto no volume de pedidos quanto no faturamento. Este padrão permite um planejamento estratégico mais apurado, incluindo a gestão de estoques, a alocação de pessoal e o planejamento de ações de marketing para impulsionar as vendas em períodos de menor atividade.

## Performance Semanal

A distribuição de atividades ao longo da semana é um indicador-chave de desempenho. A análise confirma quais dias concentram o maior volume de transações e receita, geralmente os fins de semana. Em contrapartida, os dias de menor movimento, como o início da semana, representam oportunidades claras para a criação de promoções e campanhas específicas que visem aumentar o fluxo de clientes.

## Relação entre Volume e Faturamento

A visualização combinada do volume de pedidos e do faturamento serve como um termômetro da saúde do negócio. A correlação entre as duas métricas permite identificar se o crescimento da receita é sustentado apenas pelo aumento no número de vendas ou também por um aumento no valor médio dos pedidos (ticket médio). Uma receita que cresce em um ritmo superior ao volume de pedidos indica uma valorização do ticket médio, um sinal de crescimento saudável e aumento da lucratividade.

## Perfil Demográfico

A análise demográfica da base de clientes, como a distribuição por gênero, fornece informações valiosas para a segmentação de mercado. Campanhas de marketing e a comunicação geral da marca podem ser ajustadas para dialogar de forma mais eficaz com o público predominante, potencializando o engajamento.

## 3. Pontos de Atenção

A inspeção inicial revelou alguns valores nulos/células vazias. Embora as linhas problemáticas tenham sido removidas para a análise temporal, seria importante investigar a *origem* desses dados faltantes no sistema de coleta para melhorar a qualidade dos dados no futuro. Como para esse projeto recebemos muitos dados além de anonimizados, também randomizados, consideramos que não era relevante tratá-los no momento.

## 4. Próximos passos - Derivação de Dados

Com a base de dados atual, podemos enriquecer a análise criando novas variáveis que seriam relevantes para o dashboard:

- **Ticket Médio por Cliente:** Calcular o gasto médio de cada cliente para identificar e indicar o que seriam "clientes de ouro".
- **Frequência de Compra:** Calcular o tempo médio entre os pedidos de um mesmo cliente. Indicador de lealdade.

- **Segmentação Recência, Frequência, Valor Monetário:** Técnica de marketing para segmentar clientes:
  - **Recência:** Quando foi a última compra?
  - **Frequência:** Com que frequência ele compra?
  - **Valor Monetário:** Quanto ele gastou no total?
- **Ciclo de Vida do Cliente:** Com base nos dados de pedidos, podemos classificar os clientes como "Novos", "Ativos", "Em Risco" (não compram há um tempo) ou "Perdidos", permitindo ações proativas para retenção.