

## O case:

- O case a ser resolvido tem como contexto a realidade de uma empresa varejista
- São 3 datasets, contendo informações sobre vendas (produto, quantidade, preços) e sobre as lojas em que essas vendas foram feitas
- O objetivo é analisar todas as informações e fornecer uma previsão de vendas por loja
- Preste bastante atenção nas variáveis utilizadas. Quais delas fazem sentido para a previsão do modelo? Quais são apenas sujeira? O entendimento do negócio é fundamental para o desenvolvimento de um bom modelo. Sinta-se à vontade para criar novas variáveis também à partir dos dados existentes

## As variáveis:

### Dataset Stores

- **Store** - ID da loja
- **Type** - Tipo baseado no tamanho da loja
- **Size** - Tamanho da loja




### Dataset Features

- **Store** - ID da loja
- **Date** - Data da semana
- **Temperature** - Temperatura média na região
- **Fuel\_price** - Preço do combustível na região
- **Markdown** - Promoções/remarcações - disponível apenas depois de Nov/21
- **CPI** - Inflação no período
- **Unemployment** - Taxa de desemprego
- **isHoliday** - Se a semana é um feriado ou não

### Dataset Sales

- **Store** - ID da loja
- **Dept** - O número do departamento
- **Date** - Data da semana
- **Weekly\_sales** - Vendas por departamento em cada loja
- **isHoliday** - Se a semana é um feriado ou não

## Referências

- Como eu não sou nenhum sádico e meu objetivo é que vocês aprendam o máximo possível fazendo o case, aqui estão algumas referências para que vocês consigam desenvolver o problema, lembrando que **qualquer dúvida, podem me chamar no privado também.**
-  [Linear Regression, Clearly Explained!!!](#)
- [Supervised Learning with scikit-learn - Datacamp](#)
-  [Regression Trees, Clearly Explained!!!](#)
-  [Random Forest Regression | Python](#)
- [Data visualization](#)

## Passo a passo do case:

1. Entendimento do problema
  - a. Qual é o problema? Qual a causa (ou possível causa)?
  - b. Por que resolver?
  - c. Qual pode ser uma boa métrica para avaliação do sucesso?
2. Entendimento dos dados
  - a. Quais os dados disponíveis e que informações podemos tirar deles?
  - b. Qual a variável target?
3. Pré-processamento
  - a. Limpeza de dados
  - b. Criação de outras variáveis
  - c. Normalização
  - d. Padronização
  - e. Lidar com outliers
  - f. Valores faltantes
4. Modelagem
  - a. Seleção, treinamento e teste de um algoritmo
  - b. Tunagem de hiperparâmetros
  - c. Documentar motivo da escolha
5. Avaliação
  - a. Analisar as métricas
  - b. Formular hipóteses de como o seu modelo poderia melhorar e iterar (documentar essas hipóteses)

### **Prazos:**

- Vamos trabalhar com entregas intermediárias
- Partindo do passo a passo sugerido, a ideia é que vocês entreguem pelo menos até a análise exploratória dos dados (etapa 3) no dia **25/03** e apresentem na reunião de IA
- O case completo será entregue e apresentado no dia **01/04**
- Aproveitem a entrega parcial na reunião de IA para tirarem dúvidas
- Vocês podem alterar e melhorar a análise feita na primeira entrega antes de entregar a versão final
- Por fim, espero que vocês se divirtam muito, pesquisem muito e perguntem muito. Em um projeto de data science, é sempre mais importante se fazer boas perguntas do que ter as melhores respostas.