

# A questão de negócio:

1.1) A meta principal do projeto é através do insights que teremos, descobrir qual é a melhor loja da rede Walmart para receber um investimento de expansão.

1.2) Podemos entender qual a melhor loja para receber a expansão através de primeiramente um tratamento de dados para entender quais dados são relevantes para nossa análise, após isso conseguimos através dos insights, padrões encontrados em um número de lojas que as levou a ter um sucesso.

# Entendimento do negócio

## 2.1) Quais dados eu tenho disponível?

Temos diversos dados disponíveis como por exemplo: venda na semana, a data da semana, a loja correspondente, a taxa de demissão e também temos as temperaturas que também podem influenciar no desempenho da loja.

## 2.2) Eles são relevantes para o problema?

Eliminamos alguns dados que não eram relevantes como Fuel\_price e CPI

## 2.3) Eles me trazem uma solução direta?

Maioria dos dados não tem tanta relação assim quanto ao nosso objetivo de saber qual o melhor investimento, porém com o histórico de produtividade de cada loja a gente conseguiu fazer nossa análise.

## 2.4) Qual o meu setor ou range de negócio?

O Setor é a área comercial, então nos baseamos nos históricos de vendas, as médias e somatórias para entender qual loja tem a melhor produtividade de conversão de vendas.

# A coleta de dados

3.1) Os dados que eu tenho fazem sentido?

Os dados fazem sentido, porém seria interessante também saber o custo de cada loja e também o número de funcionários, para que estivéssemos mais parâmetros internos.

3.2) Estão no formato que eu gostaria?

Sim, não houve nenhum tratamento do formato dos dados, pois os dados estavam bem organizados e prontos para a análise

3.3) O que mais eu consigo obter de informação destes dados?

Conseguimos entender também qual a relação de temperatura com o desempenho das lojas e a mesma coisa pode ser feita com semana com/sem feriado, de forma que nós buscamos um padrão para poder mensurar como esses fatores influenciam na produtividade.

# Tratamento dos dados

## 4.1) Remoção de dados outliers e valores ausentes

```
walmart = pd.read_csv('Walmart.csv')
walmart.drop(['Fuel Price', 'CPI'], inplace=True, axis=1)
walmart
walmart = walmart.dropna(axis = 0, how = 'any')
walmart['Date'] = walmart['Date'].replace('-', '', regex=True)
walmart
np.where(walmart['Temperature'] < 5)
walmart.drop([910], axis=0, inplace=True)
```

# Exploração de dados

5.1) As métricas importantes é a média de venda de cada loja, a somatória, as semanas, feriados e temperaturas

5.2) A melhor loja para investimento é a loja 20.

