



DNC.Group
Rafael da Silva Oliveira #37828
Deliverable 02 - Slot 04
Mini Projeto Walmart

link para o Colab utilizado:

<https://colab.research.google.com/drive/1r3kzt1ymSb6YuJ4vqrO-ib3lqTlr-dQx#scrollTo=hplpYhmluUC8>

link para o github:

<https://github.com/Rafa516/Walmart-Inc>

1. Introdução

Walmart, Inc., é uma multinacional estadunidense de lojas de departamento, a companhia foi fundada por Sam Walton em 1962, incorporada em 31 de outubro de 1969 e feita capital aberto na New York Stock Exchange, em 1972.

No ano de 2021, obteve um lucro de \$13.51 Bilhões.

Sendo uma das principais lojas de varejo do mundo, os dados contemplam as vendas semanais de 45 lojas espalhadas pelos Estados Unidos.

O Walmart realiza vários eventos promocionais de descontos ao longo do ano.

Essas remarcações precedem feriados importantes, os quatro maiores de todos, que são o Super Bowl, o Dia do Trabalho, o Dia de Ação de Graças e o Natal. As semanas que incluem esses feriados tem um peso maior.

2. Business Understanding

2.1 Objetivos de Negócio

A empresa Walmart deseja fazer um investimento em uma das lojas localizadas nos Estados Unidos a fim de expandir seu tamanho, foram avaliados os dados de 45 lojas, com as seguintes informações:

- Semana de Venda;
- Venda Naquela Semana;
- Flag se é ou não semana com feriado (1 - Holiday Week 0 - Non-Holiday Week);
- Temperatura do dia em °F;
- Preço do combustível na região da loja;
- Índice de preços ao consumidor; e
- Taxa de desemprego.

A loja com maior faturamento somado no período foi a de nº 20, tendo sua média semanal acima da média geral em 100% das semanas e acima da própria média em 52 semanas.

Foi possível constatar também que dentre as 10 semanas que mais venderam no período, apenas 2 eram feriados e as 2 maiores eram véspera de natal. **Objetivo principal foi analisar os dados e verificar a melhor loja para receber o investimento.**

2.2 Business e Data Understanding

O conjunto de dados fornecidos, se mostraram insuficientes, embora tenha dados relevantes, que certamente contribuíram na tomada de decisão, faltaram os dados como tamanho de cada loja e seu Zipcode, assim sendo, os dados disponibilizados foram analisados através de gráficos e tabelas geradas.

Com o levantamento feito foi possível determinar:

- As Lojas com maiores médias de Vendas Semanais;
- As Lojas com Maiores Vendas em Semanas de Feriados;
- Preço médio de venda semanal por loja;
- A loja com maior venda acumulada (soma de vendas de todo o período) e em quantas semanas do ano ela ultrapassou a média do período;
- Data frame das vendas máximas, mínimas e médias de todas as lojas ao longo do tempo;

Comparativo dos preços de combustível na região da loja;

- Comparativo de índice de Preços ao Consumidor (CPI);
- Comparativo de taxa de desemprego;
- Índice de temperatura: para avaliar em quais lojas a temperatura média tem a possibilidade de nevar;
- Médias por loja e visualização e ordenação decrescente pelas maiores médias de venda

Neste contexto, foi possível tirar alguns insights e o principal deles, o de prever a loja que já está vendendo bem e há espaço para crescimento.

2.3 Avaliando a Situação

Com o objetivo de definir qual loja Walmart, daquelas que constam na base de dados fornecida, é ideal para receber investimento direcionado para sua expansão, foram realizados os seguintes processos: ranqueamentos das lojas de acordo com: suas vendas semanais médias, preço do combustível na região da loja, Índice de Preço ao Consumidor, taxa de desemprego e temperatura, assim como seus respectivos barplots para melhor visualização dos dados.

Desta forma, é possível visualizar dentre os que mais vendem, o que eles têm em comum e principalmente se, dentre estes que mais vendem, tem algum que se destaca por ter índices diferentes.

2.4 Análise dos dados

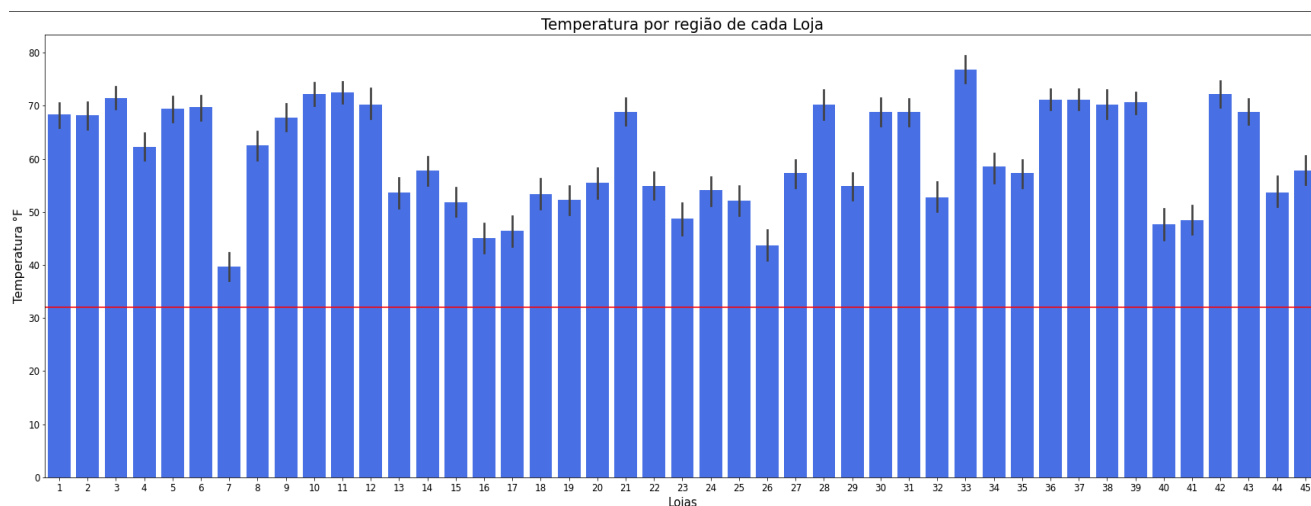
Inicialmente, foi feito um levantamento dos índices da correlação de spearman onde concluiu-se que as variáveis quantitativas (número de vendas semanais, temperatura, preço do combustível, CPI e taxa de desemprego) não apresentaram correlação alta entre si.

A variável "holiday flag", por ser uma variável categórica binária, pode ter sua correlação avaliada com outra variável quantitativa através da correlação de Pearson, mas também não observa-se uma correlação alta com o número de vendas semanais.

Foi possível também visualizar o preço médio de venda semanal por loja. Outro item importante na análise foi a criação de um Data Frame das vendas máximas, mínimas e médias de todas as lojas ao longo do tempo e o gráfico destes dados.

Além disso, os dados das colunas **Fuel_Price**, **CPI**, **Unemployment** e **Temperature** foram utilizados para comparativos a fim de conseguir informações relevantes. Fuel_Price foi levado em consideração de acordo com a região da loja, quanto mais barato, mais fácil de mais pessoas irem até a loja. O CPI, índice de preços ao consumidor, quanto menor o CPI, mais fácil os clientes comprarem mais na loja. Unemployment, quanto maior a taxa de desemprego, no mínimo, menor será o gasto das pessoas desempregadas. A coluna Temperature é uma variável importante pois a temperatura média é considerada neve, lembrando que 32 °F corresponde a 0°C.

Na figura abaixo, a linha vermelha apresenta a temperatura em que há possibilidade de nevar, o que dificulta a ida de clientes às lojas.



Resultado do Projeto

Após avaliar quais as 5 lojas que tiveram a maior média de vendas do período, foi possível ver que uma delas tinha índices melhores de taxa de desemprego, Temperatura (ficando menor parte do tempo com neve), preço de gasolina entre os mais baixos e CPI melhor.

Com base nos dados obtidos, a loja ideal para receber a expansão é a de nº **20**, pois por ter a maior média de vendas semanais, ultrapassando o valor médio de vendas semanais global em todas as semanas e também a maior venda acumulada.

