

Relatório de Análise de Vendas do Walmart utilizando a metodologia CRISP-DM

Elaborado por: Valdir Nascimento dos Santos - RDI – 97851

Data: 24 de março de 2023

Entrega 2 do Slot 4

Introdução

O Walmart é uma das maiores redes varejistas do mundo e apresenta um histórico de crescimento significativo ao longo dos anos. Este relatório tem como objetivo analisar o nível de faturamento semanal de 45 lojas do Walmart de 2010 a 2012 e apresentar sugestão de investimento em uma das suas unidades. Utilizo para isso, a metodologia CRISP-DM.

Fase de Entendimento do Negócio

O Walmart possui uma vasta gama de produtos e atende uma grande variedade de clientes. Durante o período de análise (2010 a 2012), a economia dos Estados Unidos passou por um momento de recuperação após a crise financeira global de 2008. Com isso, espera-se que o Walmart tenha experimentado um aumento nas vendas durante este período. Apresento aqui a observação da crise, pois ela deve influenciar nas escolhas que serão feitas, já que se trata de um período pós turbulência econômica.

Fase de Entendimento dos Dados

Os dados utilizados nesta análise foram fornecidos pelo Walmart e consistem em informações de vendas diárias de 45 das suas lojas nos Estados Unidos, de 2010 a 2012. Cada registro contém informações sobre a semana da venda, a loja, preço de combustível na região, semana com e sem feriado, índice de preço ao consumidor e índice de desemprego. Os dados são insuficientes para uma melhor tomada de decisão. Falta por exemplo: a região e os itens mais vendidos nestas regiões, e qual o comportamento do cliente em relação ao consumo e ao preço, quantidade de estoque e departamentos (laticínios, bebidas, higiene etc). Pesquisei algumas informações, mas como se trata de dados particulares, acabam sendo restrito e de pouco acesso. O *dataframe* foi montado a partir do arquivo disponibilizado no formato csv (walmart.csv), o mesmo tem o nome de walmart, mas vou alterando o nome de acordo com a necessidade da análise.

Fase de Preparação dos Dados

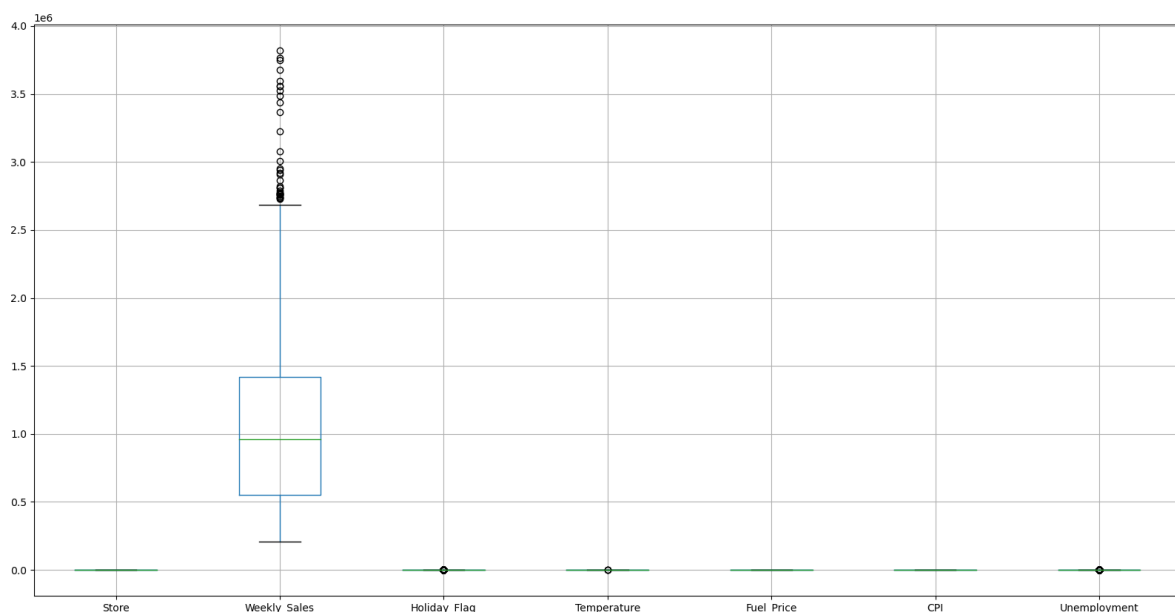
Antes de iniciar a análise, foram verificados se havia presença de outliers, registros nulos e registros duplicados e constatei que os dados estavam ok, sem necessidade de limpeza ou correções. Foram realizadas algumas transformações nos dados para facilitar a análise, a data era do tipo *objects*, transformei em *datetime* para selecionar períodos específicos. Foram criadas variáveis para representar as vendas totais de cada loja, bem como a variação percentual nas vendas em relação ao ano anterior. Criei alguns atributos temporários para o *dataframe* com a finalidade de fazer comparações mais profundas dos dados.

Fase de Modelagem e Avaliação

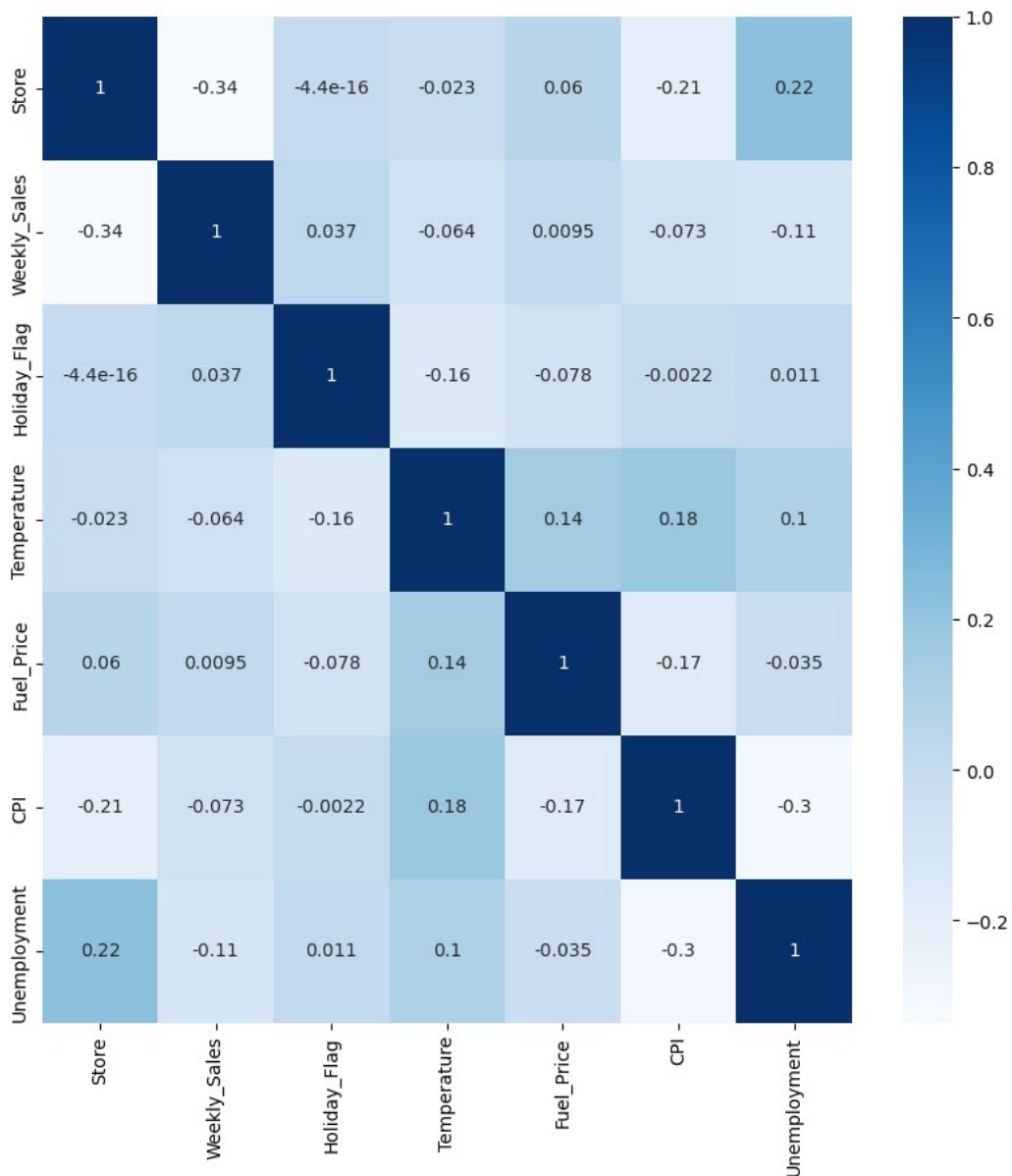
A análise foi realizada utilizando técnicas de modelagem de dados e estatísticas descritivas. Foram realizados gráficos de linha para analisar a evolução das vendas ao longo do tempo e gráficos de barras para comparar as vendas entre as lojas. Além disso, foi realizado um modelo de regressão linear para identificar quais variáveis têm maior impacto nas vendas.

Como mencionado anteriormente, este gráfico foi plotado para melhor entender a possibilidade de haver *outliers*, o qual constatei que não, são mínimas e justificáveis a distancia de certos valores para a média.

Gráfico Box - Análise de vendas do Walmart de 2010 a 2012



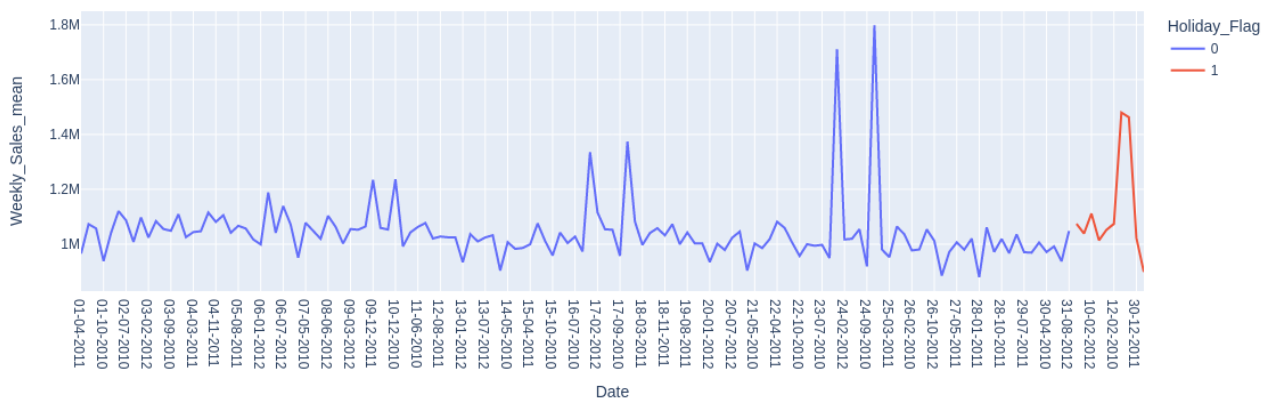
Para buscar um entendimento do por quê algumas lojas se destacam das outras, plotei um heatmap para visualizar melhor o nível de relação entre as variáveis:



- A correlação entre as vendas semanais (Weekly_Sales) e a loja (Store) é negativa, o que significa que há uma relação inversa entre essas variáveis. Neste caso, as vendas semanais tendem a diminuir à medida que o número da loja aumenta.

- Há uma correlação positiva entre a temperatura e as vendas semanais, o que indica que as vendas tendem a aumentar com o aumento da temperatura, mas não afeta tanto assim as vendas..
- O preço do combustível (Fuel_Price) e o índice de preços ao consumidor (CPI) têm correlação negativa com as vendas semanais, o que sugere que os consumidores podem gastar menos quando os preços do combustível e do CPI são mais altos. Isto me leva observar as lojas com proximidade a postos em que a gasolina seja mais barata ou a distância entre loja, posto e residência sejam mais curtas, mas não tem como afirmar com certeza, pois não dispomos destes dados.
- Há uma correlação negativa entre o índice de preços ao consumidor (CPI) e a taxa de desemprego (Unemployment), o que sugere que o aumento dos preços pode levar a uma redução da taxa de desemprego se o faturamento da loja aumentar. Mas o desemprego também diminui o poder aquisitivo do cidadão, levando a uma queda nas vendas
- A correlação entre as vendas semanais e a Holiday_Flag é baixa, o que sugere que a ocorrência de feriados não tem uma forte influência nas vendas semanais, esta afirmação será confirmada em um gráfico de linhas que demonstra claramente esta baixa correlação:

Média de vendas por semana de todas as lojas considerando o Holiday_Flag



Este gráfico sugere que nos feriados (em vermelho), houve sim um leve aumento nas vendas, o maior pico no feriado foi verificado na semana de ação de graças (Christmas). Mas a maior média foi verificada em tempo normal (sem feriado). Este outro gráfico, pode demonstrar melhor a relação venda x feriados:

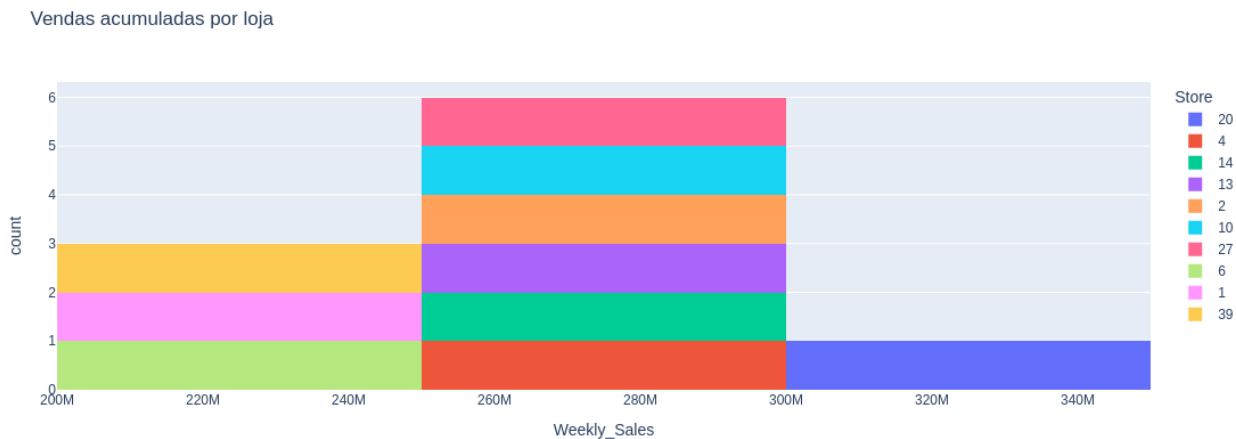


Dentro da análise já estava convencido de que a Store 20 era a loja que deveria ser investida em ampliação. Porém, ao analisar a quantidade de vendas a cima da média, percebe-se que muitas lojas têm uma venda a cima da sua própria média, em um quantitativo maior que a Store 20, embora a 20 seja mais lucrativa e tenha um rendimento estável. Busquei então fazer um comparativo do percentual da variação das vendas em relação ao ano anterior:



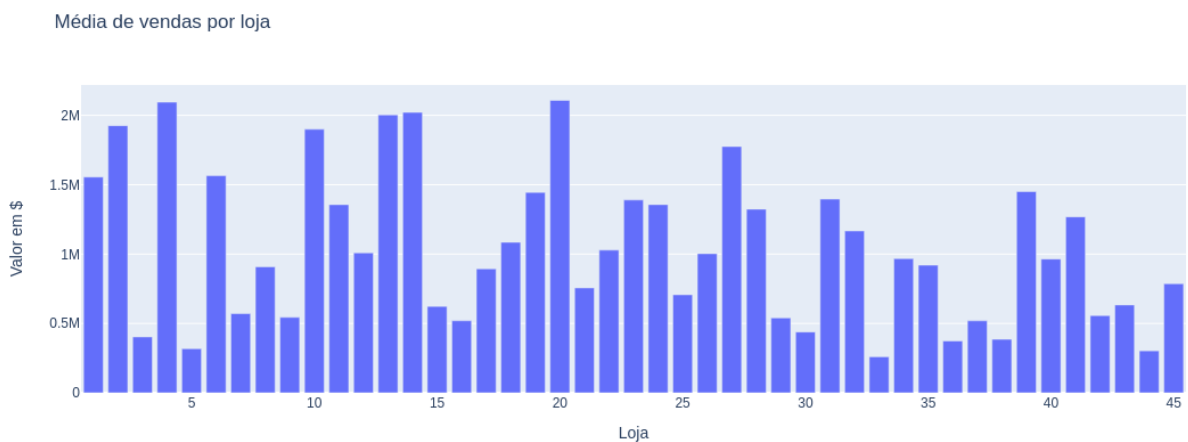
Observa-se que a Store 4 e 6 tem uma variação positiva de um ano para o outro maior que as demais.

No histogram de vendas acumuladas por loja, fiz um top 10, para analisar no gráfico de variação quais delas apresentou maior crescimento de um ano para o outro.

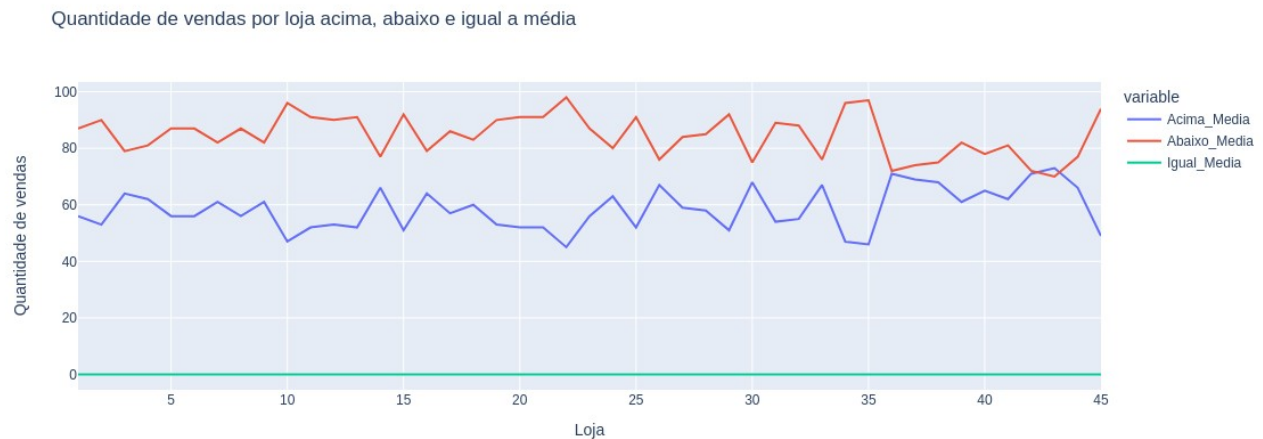


É possível verificar entre os dois gráficos, que a loja 20 tem maior venda acumulada, seguida das demais, porém, sua variação positiva é menor que a maioria. Pode-se supor que ela atingiu seu nível máximo de vendas e que as outras podem crescer ainda mais, os dados são poucos para maiores conclusões.

A média de vendas por loja corrobora com esta análise:



As Store's 2, 4, 10, 13, 14 e 20, têm uma média interessante. A que mais chama atenção no entanto é a loja 4, que tem um crescimento 4.20% (vide gráfico dinâmico de variação no google colab). Verifiquei também a frequência com que cada Store teve suas vendas acima, abaixo ou igual à própria média:



As lojas nas quais classifiquei como top 10, apresentaram um percentual interessantes de semanas com vendas a cima da média ou abaixo dela. Dentre elas se destacaram mais uma vez as lojas 4, 13 e 14, a 20 permaneceu estável e teve a média abaixo mais vezes, porém sua média é muito alta, mas a loja 4 também tem uma média alta e superou a própria média mais vezes.

Conclusão

Os resultados da análise mostraram que as vendas do Walmart cresceram de forma discreta se analisado loja a loja, mas se comparado com um período de pós recessão e o faturamento total de todas as lojas somadas, o crescimento foi interessante no período de 2010 a 2012. A loja com **melhor desempenho acumulado é a loja 20**, mas a melhor loja para se investir no momento, **dado ao seu crescimento nos lucros (4,20%)**, é a **Store 4**, pois a mesma demonstrou que ainda tem espaço para crescer.

Mas detalhes desta análise e outros gráficos e modelagem, podem ser visto no google colabatory, **neste link**:

<https://colab.research.google.com/drive/1WafqZHlnkp-gFOC-agZPGIKeLLrx7-yA?usp=sharing>