



Deliverable:

Projeto – Walmart

Willian Kakuta

RID: 29018

Turma: DS05

Sapucaia do Sul – Setembro de 2022

Descrição:

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 têm um peso maior.

Questão de negócio:

O objetivo é definir qual loja seria melhor para expandir o seu tamanho.

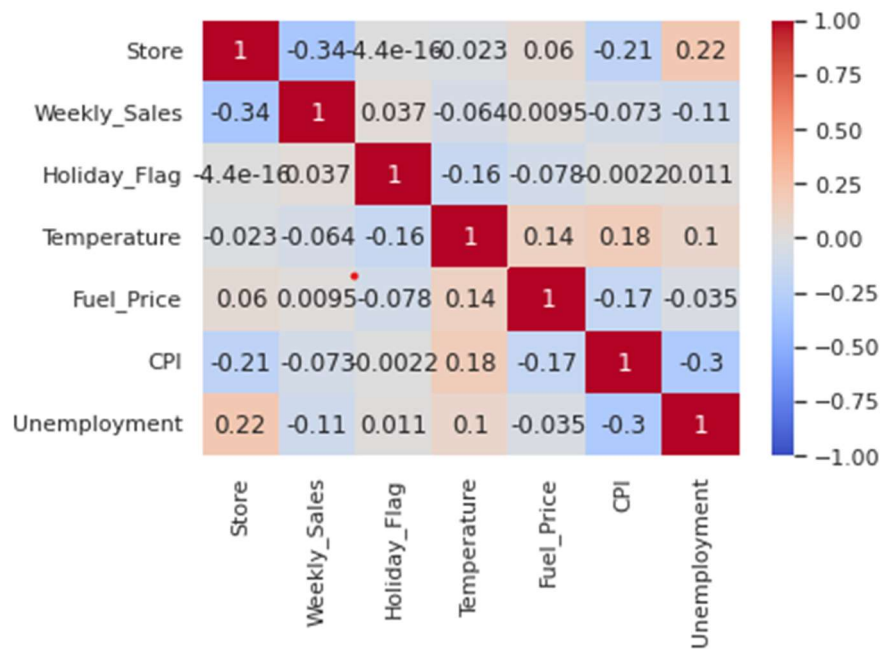
Analisando o dataset apresentado pela empresa, tratando os dados, tirando insights para o desenvolvimento de um plano de ação.

Entendimento do negócio:

Os dados que tenho disponíveis são:

- Store - Número da Loja
- Date - Semana de Venda
- Weekly_Sales - Vendas da Semana
- Holiday_Flag - Semana com Feriados (1), Semana sem Feriado(0)
- Temperature - Temperatura do Dia em °F
- Fuel_Price - Preço do Combustível na Região da Loja
- CPI - Índice de Preço ao Consumidor
- Unemployment - Taxa de Desemprego

Construí uma matriz de correlação, mapa de calor, para ver se existia alguma correlação direta entre os dados, mas não foi encontrado nenhuma correlação significativa que pudesse nos trazer uma solução direta.

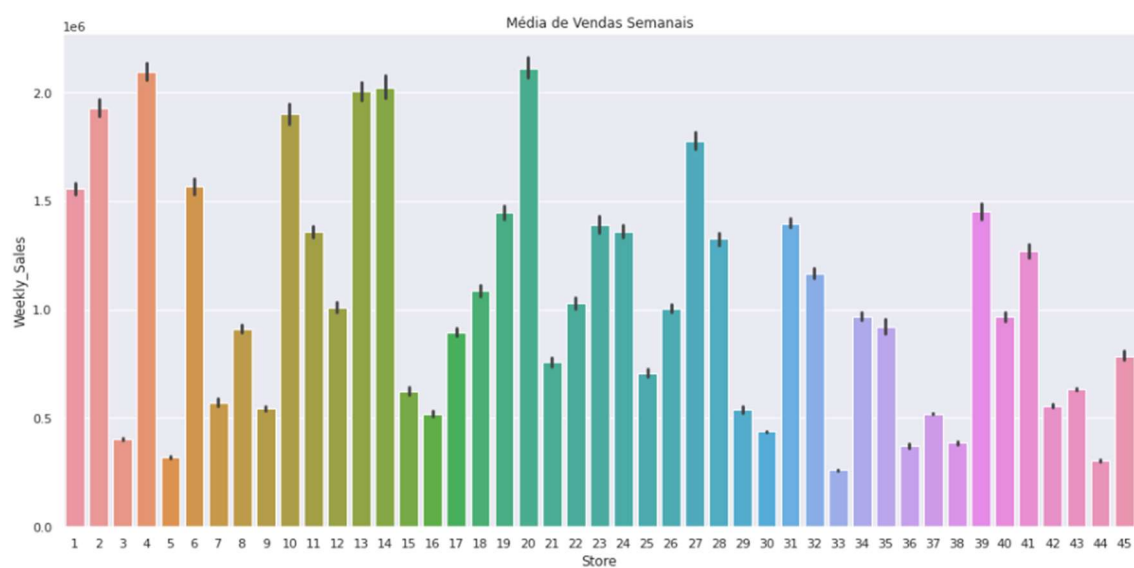


Coleta de dados:

Como não há uma correlação significativa entre os dados, o único dado que faz sentido a ser trabalhado neste momento é as Vendas da Semana.

Os dados estão em um formato bom de se trabalhar.

Calculando a média de vendas por loja conseguimos encontrar as lojas que mais vendem.

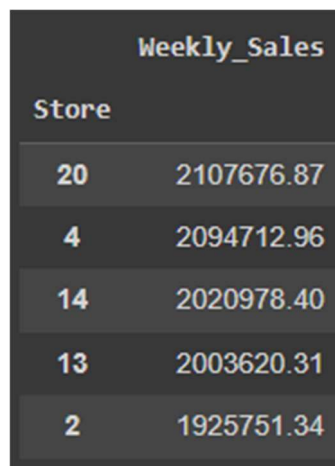


Limpeza de dados:

Não existia nenhum dado nulo. Como estamos lidando com números de vendas, não acho que faz sentido excluir os outliers, pois algum comportamento fora da curva como as vendas durante um feriado pode ser perdido. E analisaremos as outras lojas para ver se existe a necessidade de excluir algum outlier.

Exploração de Dados:

Primeiramente, fizemos a média de venda de cada loja para identificar as lojas que mais vendem.



A screenshot of a table with a dark background and light-colored text. The table has two columns: 'Store' and 'Weekly_Sales'. The data is as follows:

Store	Weekly_Sales
20	2107676.87
4	2094712.96
14	2020978.40
13	2003620.31
2	1925751.34

Depois de identificar as lojas, criei um mapa de calor para ver se existia alguma correlação entre os dados para observar se algum dos outros dados influenciaria nos resultados superiores dessas lojas, mas não houve nenhuma correlação significativa.

Então observei os dados de cada loja separadamente, vendo qual das lojas teve um desempenho melhor sobre a sua própria média. Onde se destacaram as lojas 20, 4 e 14 que por sinal são as lojas com as maiores médias e as lojas que mais venderam.

Analisei a influência dos feriados sobre as vendas e identifiquei que há um pico de vendas na semana de Ação de Graças, e um pico maior ainda na semana antes do Natal, tendo uma queda brusca abaixo da média de vendas na semana do ano novo. Mas essa informação não é algo que nos ajuda com o nosso objetivo, pois as lojas analisadas apresentam o mesmo comportamento. Então não conseguimos extrair algo relevante para o nosso objetivo dessa informação.

Analisando separadamente as lojas, como os dados relacionados as vendas são muito parecidos entre as lojas, tentei buscar os outros dados do dataset para ver se conseguia ver alguma diferença que fosse relevante. E notei que o CPI, o índice de preços ao

consumidor da loja 4 é muito menor que das outras lojas. Assim como a gasolina também, apesar de não ter uma diferença tão elevada. Mas a Taxa de Desemprego durante o tempo era de 8% foi para menos de 4%, baixando a metade, enquanto as outras lojas se mantiveram, praticamente no mesmo número durante o tempo.

Baseado nestas informações, a loja escolhida para a expansão é a loja 4. Porque o lugar que a loja se encontra, está com uma economia mais consolidada e em desenvolvimento, vendo os dados de inflação e desemprego. Além de que, a loja já apresenta uma excelente performance, diminuindo os riscos da ação.

Link Colab:

<https://colab.research.google.com/drive/120BDzDbRvdNtxLvc0VpJzy16nIONBjxZ?usp=sharing>