

RELATÓRIO WALMART

2010 - 2012



Prepared By:
Mikaely Pedrosa



A QUESTÃO DE NEGÓCIO

Objetivo: Expansão de loja

Métodos: Por meio de linguagem de programação Python, será feito um levantamento do faturamento das 45 lojas nos EUA, procurando responder as seguintes questões, e a partir das respostas, selecionar uma loja para a expansão:

1. Qual o valor médio de venda semanal por loja?
2. Qual o atributo que tem maior correlação (positiva ou negativa) com o número de vendas?
3. Da loja com maior venda acumulada (soma de vendas de todo o período), quantas semanas do ano ele ultrapassou a média do período?
4. Como estão distribuídas as vendas máximas, mínimas e médias de todas as lojas ao longo do tempo?
5. Os eventos e feriados interferem nessas medidas?
6. Das lojas com mais vendas, alguma se manteve constante durante os três anos?



ENTENDIMENTO DO NEGÓCIO

A empresa

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.

(Wikipedia)

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.

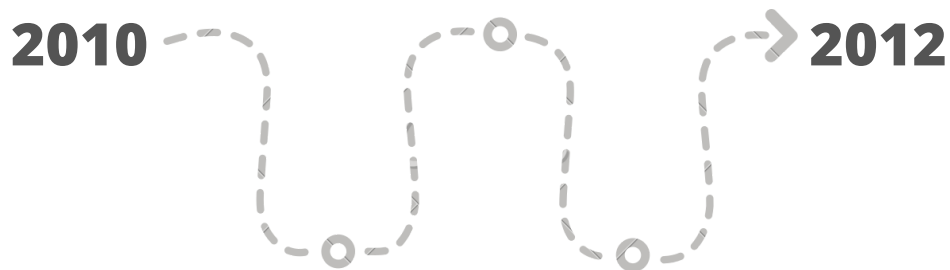
Essas remarcações precedem feriados importantes, os quatro maiores de todos, que são o Super Bowl, o Dia do Trabalho, o Dia de Ação de Graças e o Natal.

As semanas que incluem esses feriados têm um peso maior.



Dados disponíveis

Os dados apresentam as vendas semanais de 05/02/10 até 01/11/12 em **45 lojas** varejistas da rede Walmart e também algumas métricas econômicas e meteorológicas por semana.



A seguir, algumas métricas presentes nos dados que poderão ser importantes para a análise:



DATAS DE EVENTOS E FERIADOS

Super Bowl

12-Feb-10, 11-Feb-11, 10-Feb-12

Labour Day

10-Sep-10, 09-Sep-11, 07-Sep-12

Thanksgiving

26-Nov-10, 25-Nov-11, 23-Nov-12

Christmas

31-Dec-10, 30-Dec-11, 28-Dec-12



DESCRIÇÃO DAS COLUNAS



Store = Número da Loja

Date = Semana de Venda

Weekly_Sales = Venda Naquela Semana

Holiday_Flag = se é ou não semana com feriado (1 - Holiday Week 0 - Non-Holiday Week)

Temperature = Temperatura do dia em °F

Fuel_Price = Preço do combustível na região da loja

CPI = Índice de preços ao consumidor

Unemployment = Taxa de desemprego



EXPLORANDO E TRATANDO OS DADOS

Importante: caso seja necessário consultar o código de programação, o link estará disponível ao fim deste relatório.


Importando os nossos dados e visualizando as primeiras linhas da tabela, temos:

```
df = pd.read_csv('/content/walmart.csv')  
df.head(10)
```

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
0	1	05-02-2010	1643690.90	0	42.31	2.572	211.096358	8.106
1	1	12-02-2010	1641957.44	1	38.51	2.548	211.242170	8.106
2	1	19-02-2010	1611968.17	0	39.93	2.514	211.289143	8.106
3	1	26-02-2010	1409727.59	0	46.63	2.561	211.319643	8.106
4	1	05-03-2010	1554806.68	0	46.50	2.625	211.350143	8.106
5	1	12-03-2010	1439541.59	0	57.79	2.667	211.380643	8.106
6	1	19-03-2010	1472515.79	0	54.58	2.720	211.215635	8.106
7	1	26-03-2010	1404429.92	0	51.45	2.732	211.018042	8.106
8	1	02-04-2010	1594968.28	0	62.27	2.719	210.820450	7.808
9	1	09-04-2010	1545418.53	0	65.86	2.770	210.622857	7.808

Antes de tudo, tivemos alguns cuidados com os dados que recebemos, tais como:

- **Verificação de linhas nulas** -> não há linhas nulas
- **Verificação de linhas duplicadas** -> não há linhas duplicadas
- **Tratamento da coluna 'Date'** -> dia/mês/ano
- **Contagem da quantidade de Lojas** -> somam 45 lojas
- **Contagem das 'Holiday Flag'** -> '0': 5985 // '1': 450




Continuando com a exploração dos dados,
aqui temos um **resumo estatístico**:

Nele mostra a contagem, a média, o desvio padrão, a mínima, máxima e os quartis de cada coluna.

Ainda veremos essas métricas com mais detalhes.

```
[37] df.describe()
```

	Store	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment	Month	Year
count	6435.000000	6.435000e+03	6435.000000	6435.000000	6435.000000	6435.000000	6435.000000	6435.000000	6435.000000
mean	23.000000	1.046965e+06	0.069930	60.663782	3.358607	171.578394	7.999151	6.475524	2010.965035
std	12.988182	5.643666e+05	0.255049	18.444933	0.459020	39.356712	1.875885	3.321797	0.797019
min	1.000000	2.099862e+05	0.000000	-2.060000	2.472000	126.064000	3.879000	1.000000	2010.000000
25%	12.000000	5.533501e+05	0.000000	47.460000	2.933000	131.735000	6.891000	4.000000	2010.000000
50%	23.000000	9.607460e+05	0.000000	62.670000	3.445000	182.616521	7.874000	6.000000	2011.000000
75%	34.000000	1.420159e+06	0.000000	74.940000	3.735000	212.743293	8.622000	9.000000	2012.000000
max	45.000000	3.818686e+06	1.000000	100.140000	4.468000	227.232807	14.313000	12.000000	2012.000000



Para a escolha da loja que terá sua expansão,
iniciaremos respondendo as questões citadas em
'Questão do negócio'.

QUESTÃO 01.

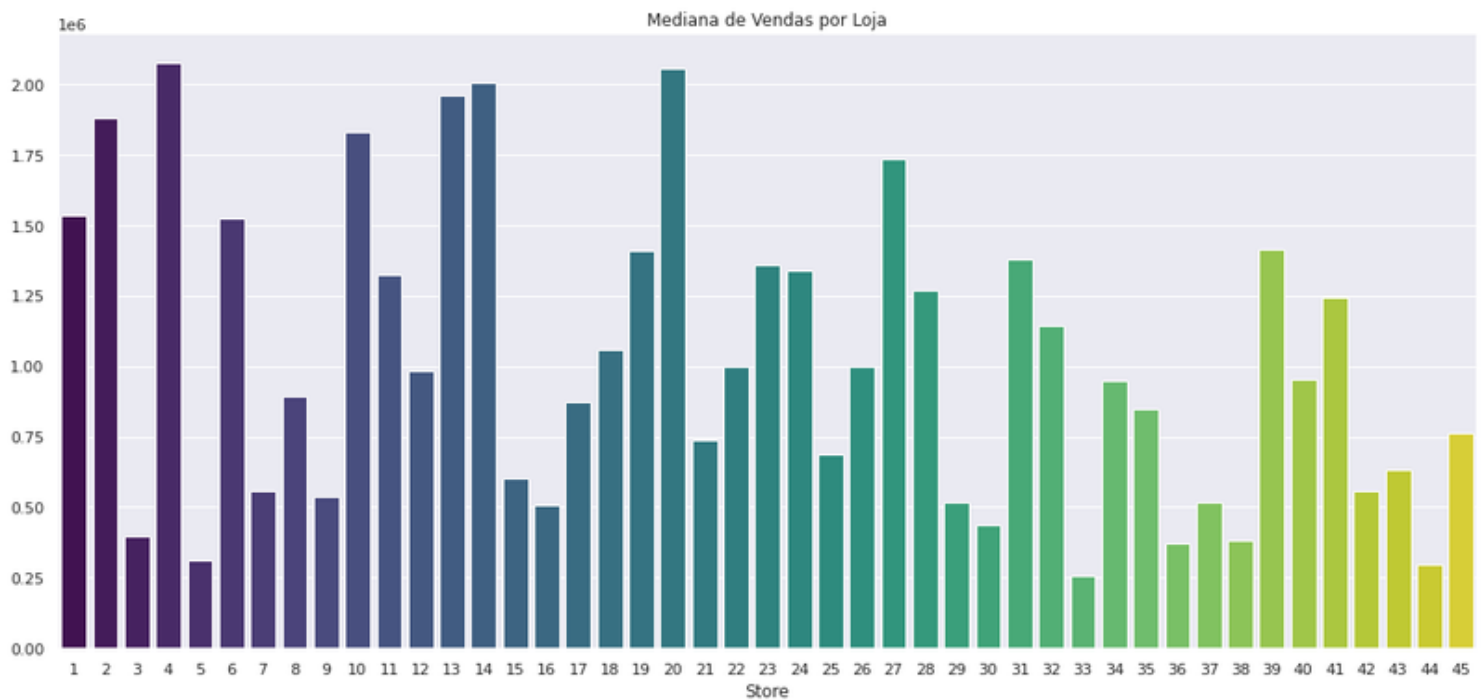
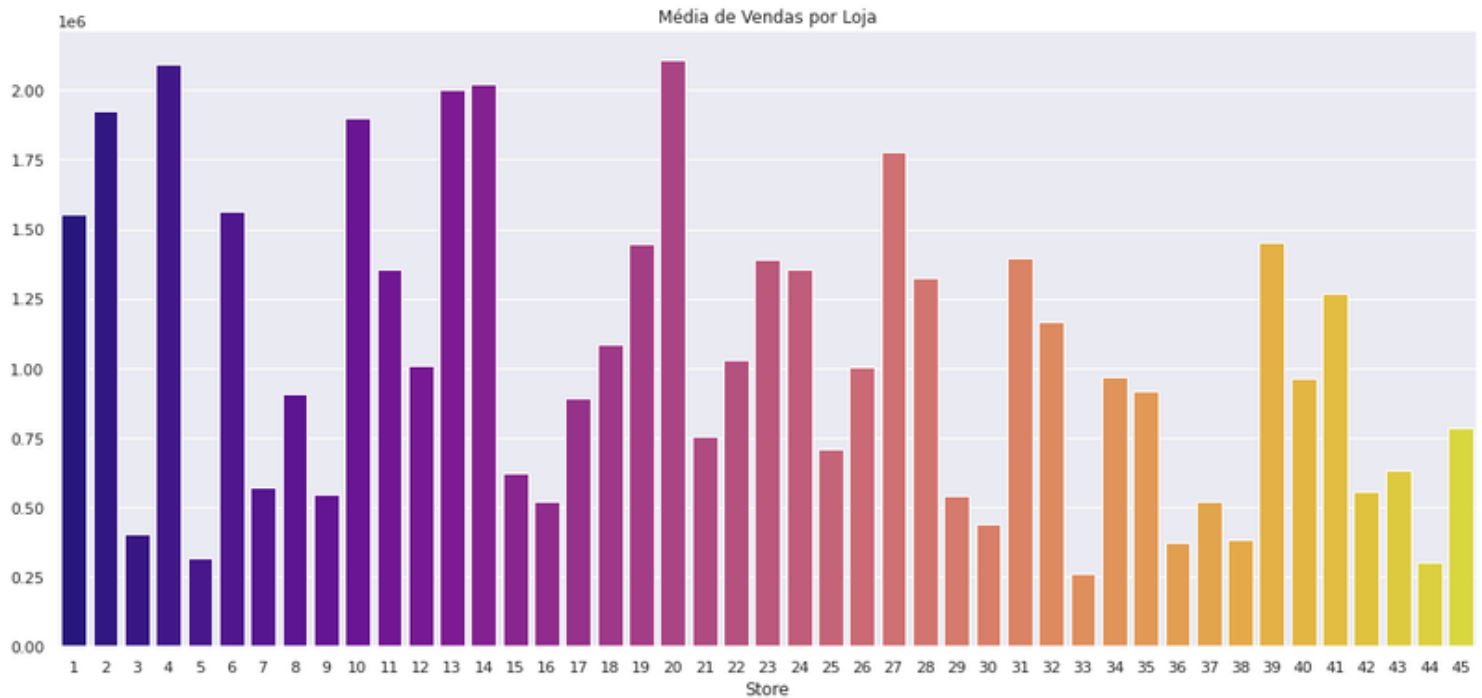
QUAL O VALOR MÉDIO DE VENDA SEMANAL POR LOJA?

Aqui veremos tanto a **média** quanto a **mediana** do valor
arrecadado por loja, ordenados do maior para o menor.
Para melhor visualização, colocaremos apenas o **top 5**:

Loja (Store)	Média de Vendas	Loja (Store)	Mediana de Vendas
20	2107676	4	2073951
4	2094712	20	2053165
14	2020978	14	2004330
13	2003620	13	1958823
2	1925751	2	1879107

Se comparadas as médias e as medianas, observa-se pouca diferença no Ranking de venda entre uma tabela e outra.

Visualmente, os gráficos se comportam muito semelhantes também, em todas as lojas:



No quesito média e mediana de vendas, já conseguimos identificar as lojas com maior faturamento, são elas:
lojas 20, 4, 14, 13 e 2.

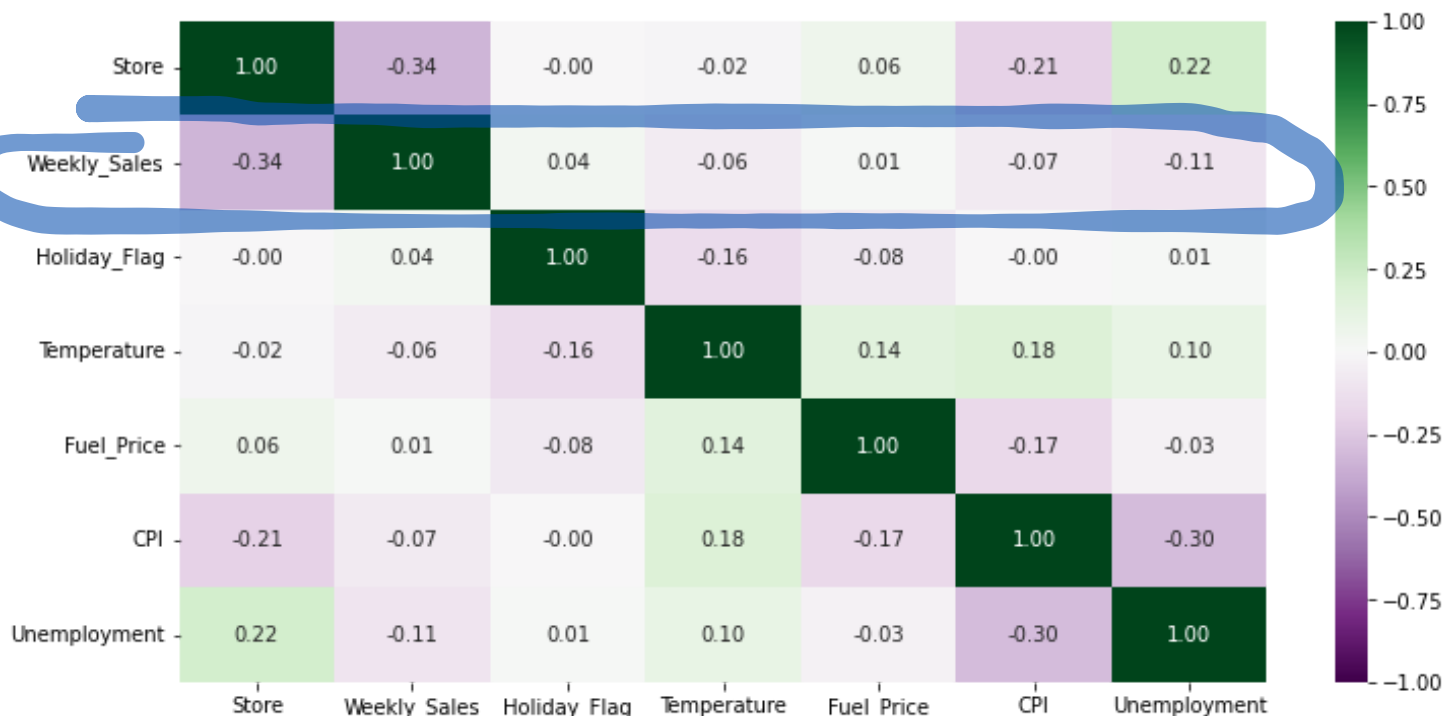
Continuaremos a análise respondendo a seguinte questão:

QUESTÃO 02.

QUAL O ATRIBUTO QUE TEM MAIOR CORRELAÇÃO (POSITIVA OU NEGATIVA) COM O NÚMERO DE VENDAS?

Abaixo temos um mapa de correlação.
Valores próximos a 1 e a -1 indicam possível correlação, positiva e negativa, respectivamente.

Vamos focar nos fatores que se correlacionam à coluna 'Weekly_Sales'



Não foi possível observar nenhuma correlação forte com a coluna Weekly_Sales, pois não há valores próximos o suficiente de 1.0 ou -1.0.

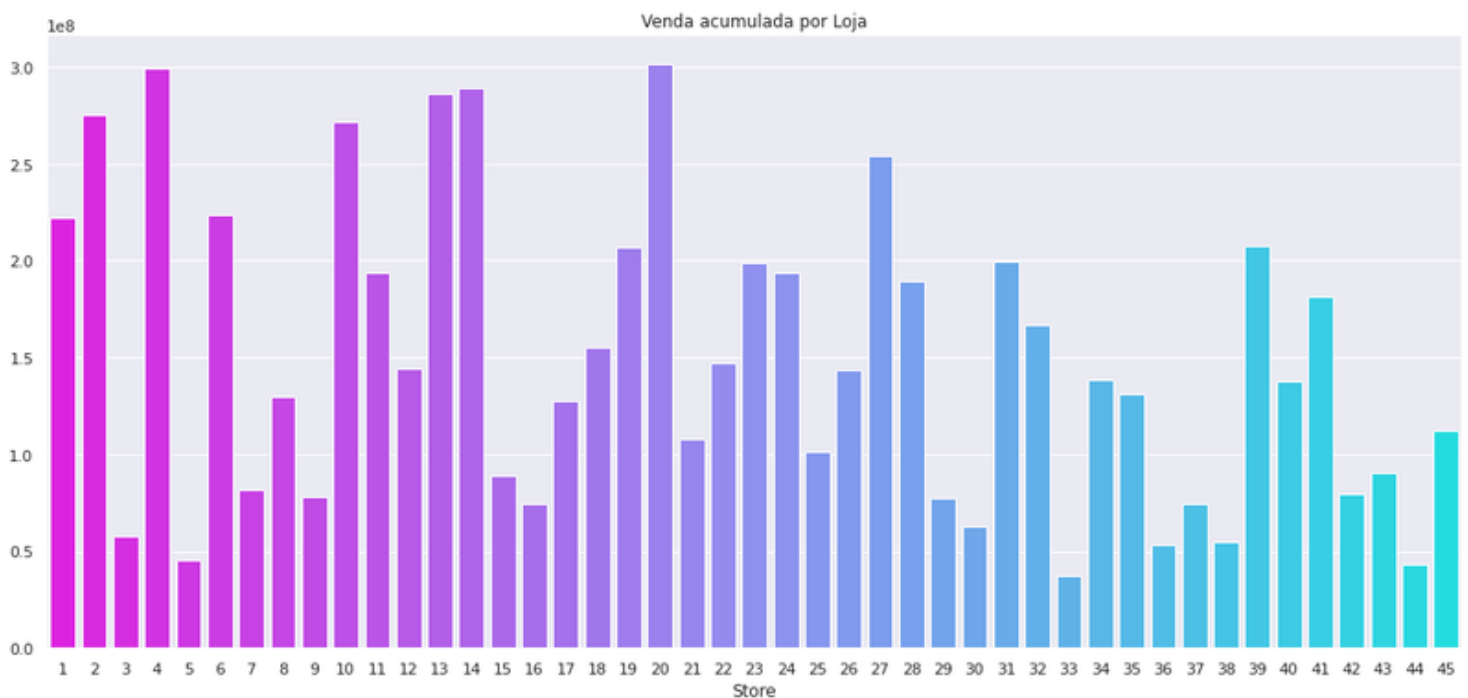
QUESTÃO 03.

DA LOJA COM MAIOR VENDA ACUMULADA (SOMA DE VENDAS DE TODO O PERÍODO), QUANTAS SEMANAS DO ANO ELE ULTRAPASSOU A MÉDIA DO PERÍODO?

O primeiro passo foi somar todas as vendas obtidas por loja durante todo o período e como resultado das top 5 lojas tivemos:

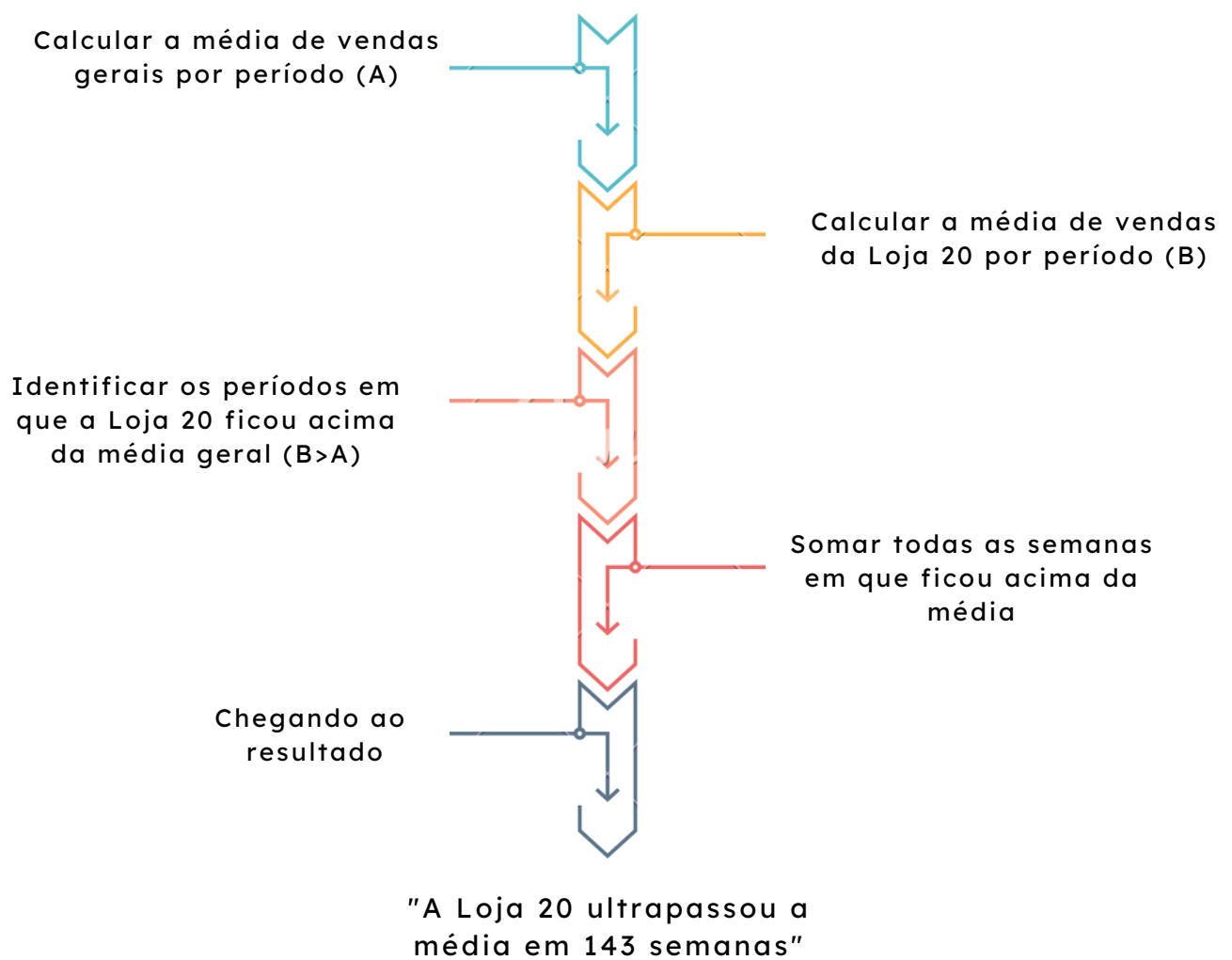
Loja (Store)	Total Vendas Acumuladas
20	301397792.46
4	299543953.38
14	288999911.34
13	286517703.8
2	275382440.98

No gráfico podemos visualizar melhor a distribuição do total de vendas por todas as 45 lojas:



Tendo em vista que a loja com o maior acumulado de vendas foi a **Loja 20**, iremos agora determinar quantas semanas do ano ela ultrapassou a média do período.

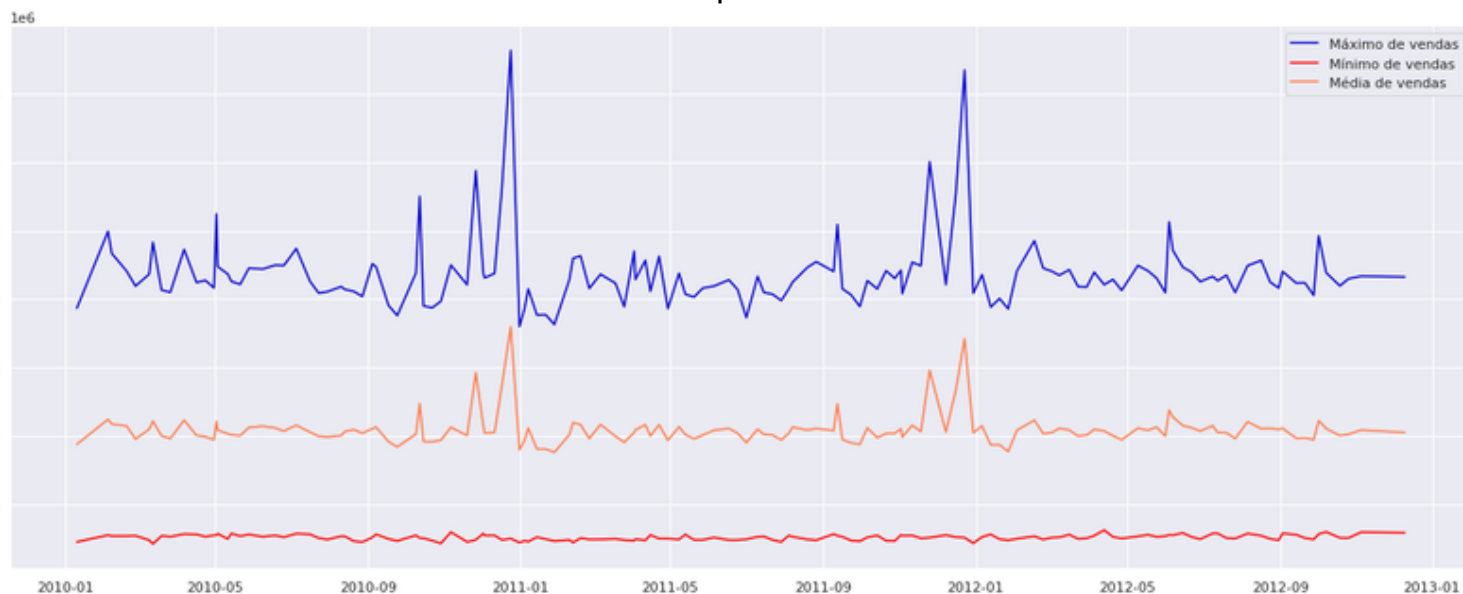
Para isso, seguimos o passo a passo:



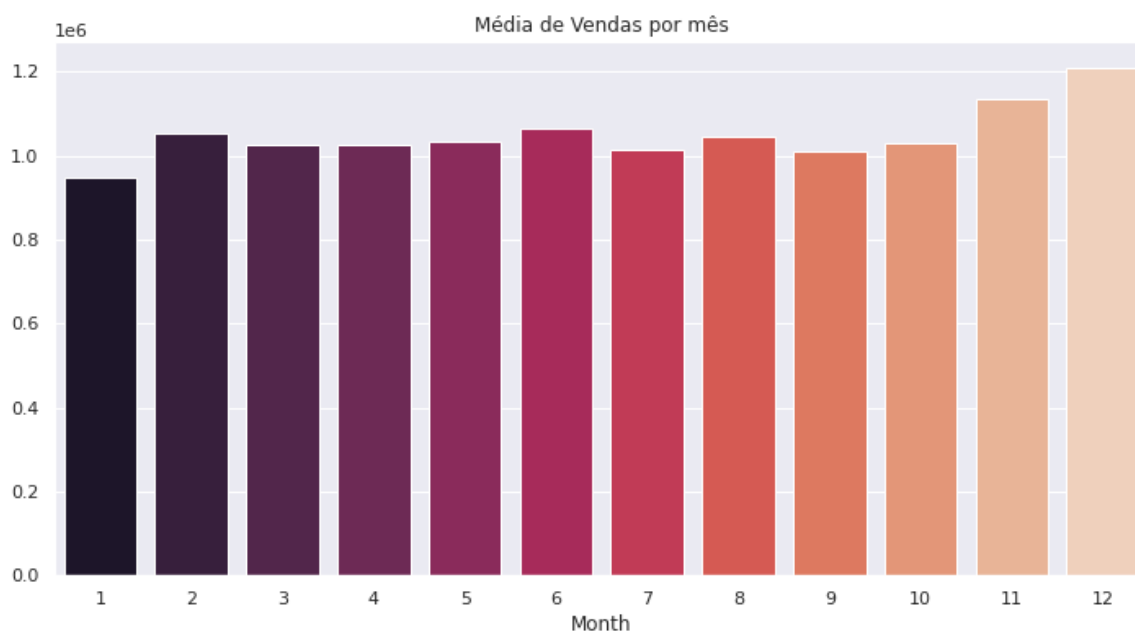
QUESTÃO 04.

COMO ESTÃO DISTRIBUÍDAS AS VENDAS MÁXIMAS, MÍNIMAS E MÉDIAS DE TODAS AS LOJAS AO LONGO DO TEMPO?

Neste gráfico temos as **máximas**, **médias** e **mínimas** das vendas durante todo o período.



Explorando mais a fundo, vemos que os meses de Novembro e Dezembro são os que possuem maior média de vendas:



QUESTÃO 05.

OS EVENTOS E FERIADOS INTERFEREM NESSAS MEDIDAS?

Adicionamos ao gráfico anterior as marcas relacionadas aos feriados e eventos:



Vemos então que o pico das vendas se concentra em torno das semanas de Thanksgiving e Christmas.

QUESTÃO 06.

DAS LOJAS COM MAIS VENDAS, ALGUMA SE MANTEVE CONSTANTE DURANTE OS TRÊS ANOS? HOUVE CRESCIMENTO?

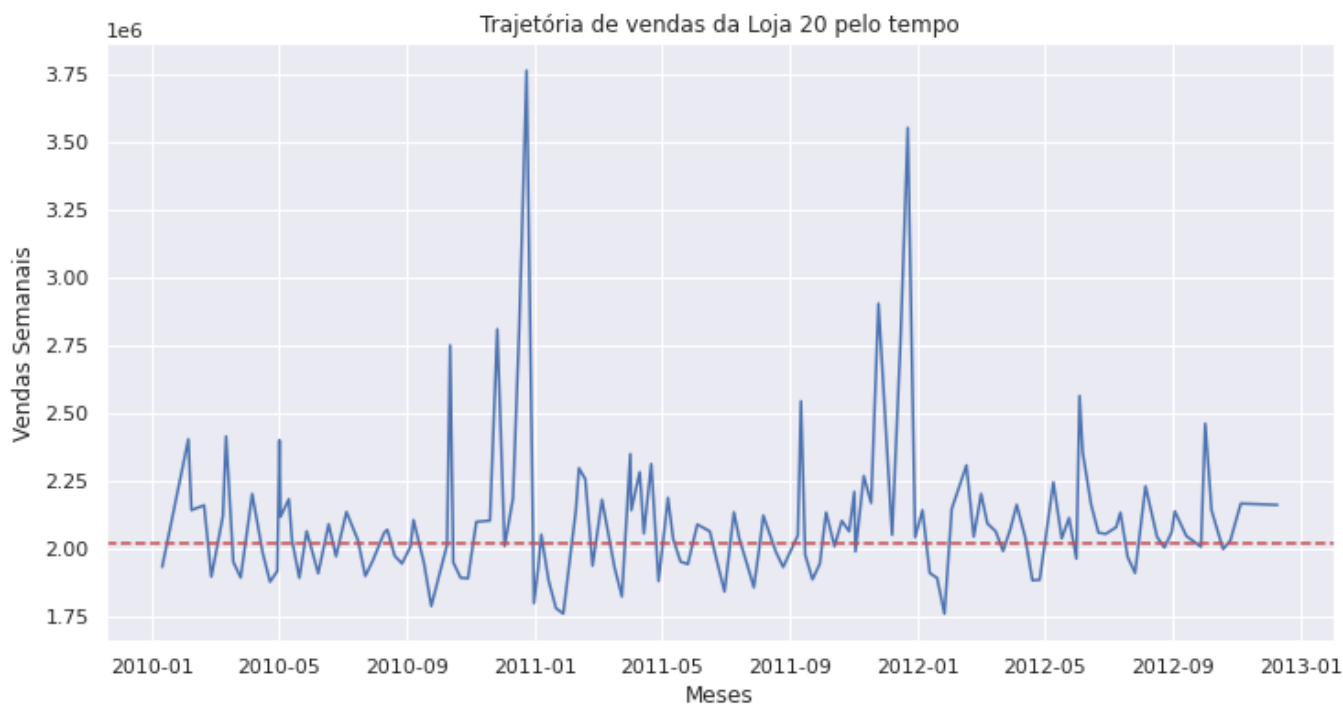
Para a expansão, iremos primeiramente considerar as lojas que possuem uma maior quantidade de vendas.

Outro ponto a ser considerado, será a constância que a loja se mantém acima da média durante todo o período, não apenas nas datas de pico, bem como se está havendo um crescimento.

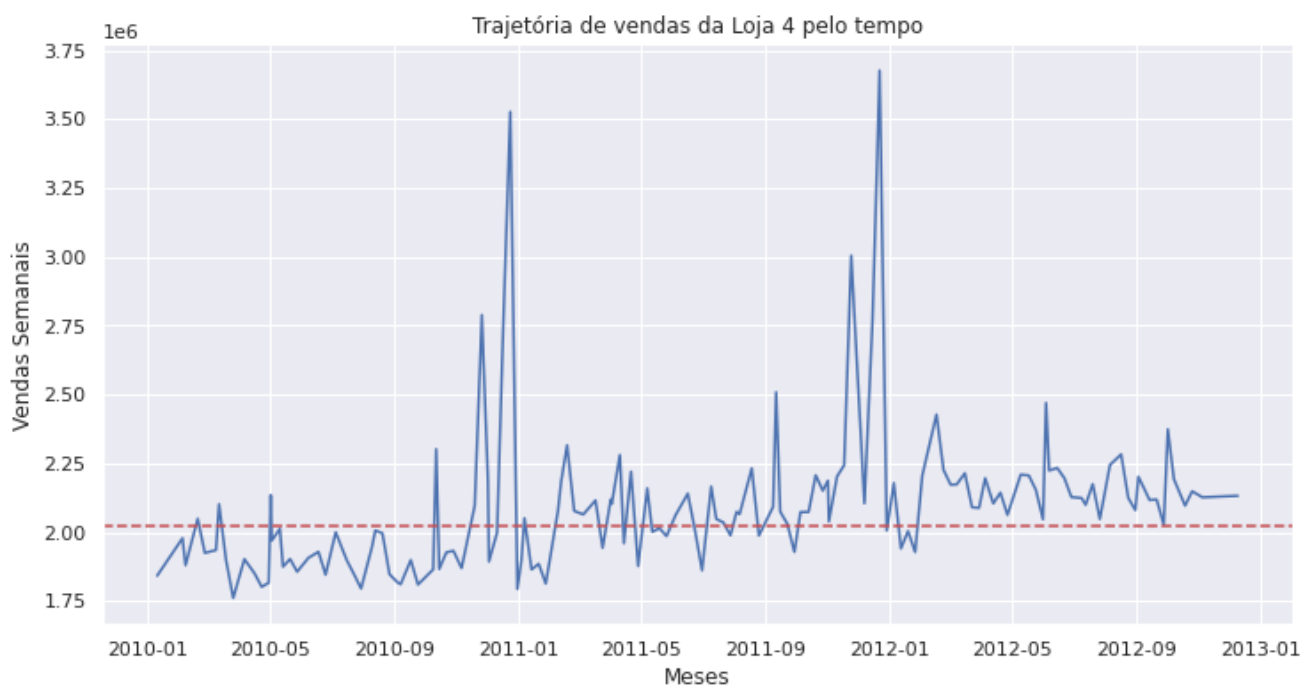
Para isso iremos comparar o desempenho das **5 lojas com mais vendas**, observando sua constância durante todo o período.

Nos gráficos abaixo, a **linha vermelha** indica a média das vendas por todo o período, observe que todas as lojas possuem picos parecidos. O que iremos analisar agora são os períodos para além dos picos.

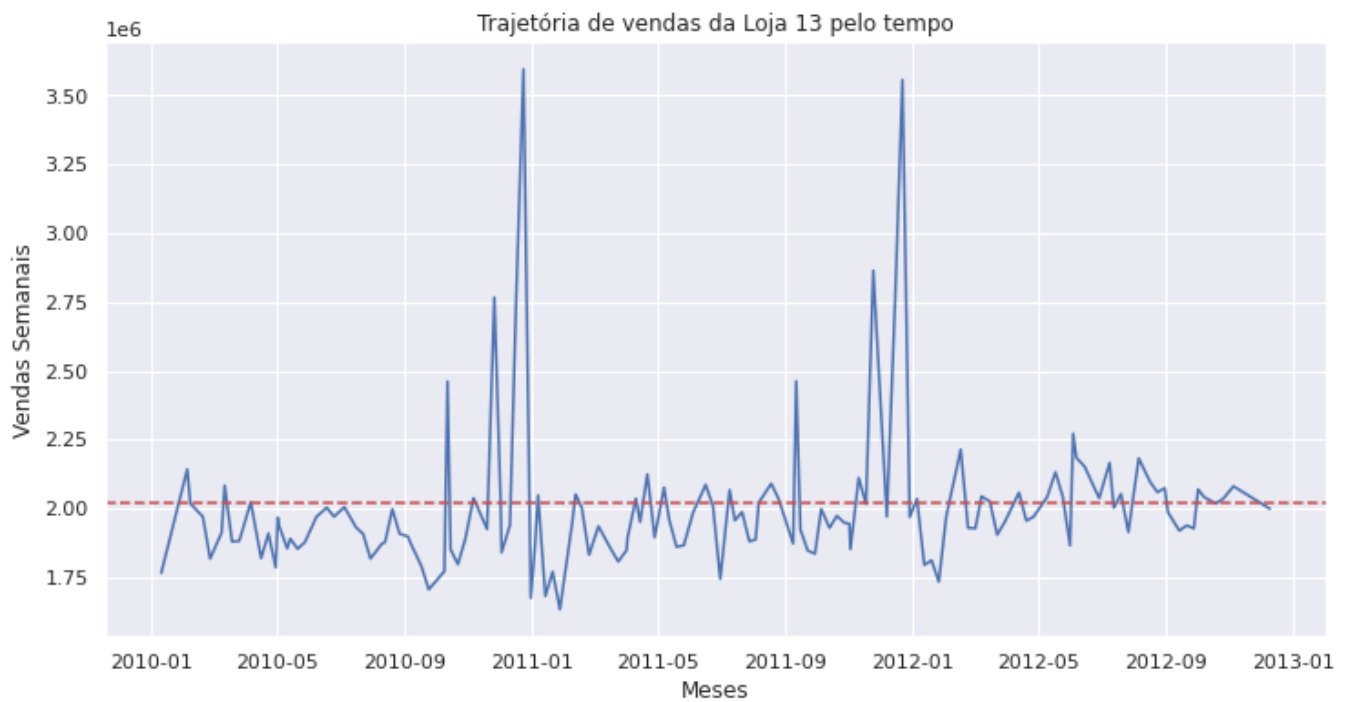
Loja 20: se mantém próxima à média, com oscilações nos anos de 2010 e 2011. Vemos um crescimento em 2012.



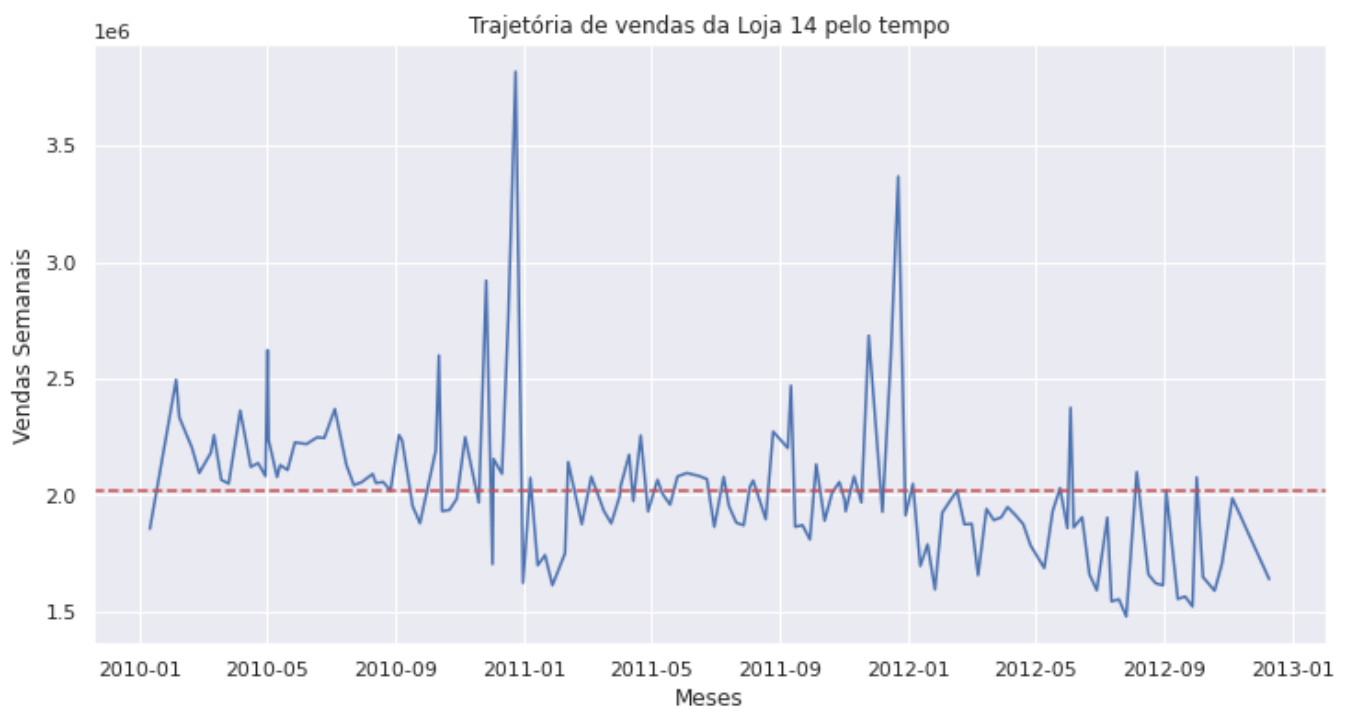
Loja 04: começa abaixo da média e possui um crescimento ordenado visível.



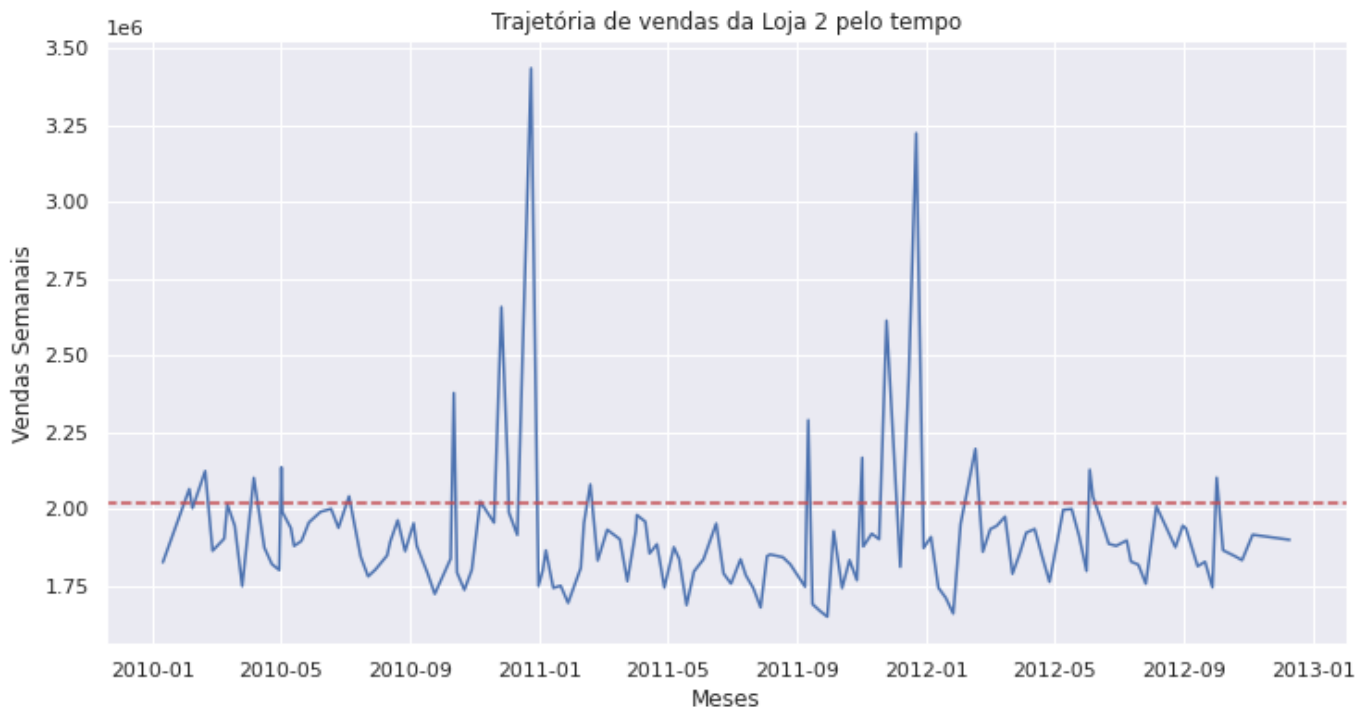
Loja 13: começa abaixo da média em 2010 e 2011, com crescimento em 2012.



Loja 14: vemos que inicia acima da média, porém apresenta queda nas vendas no último ano de 2012.



Loja 02: praticamente só fica acima da média em épocas de pico.



ESCOLHA DA LOJA

Com tudo que foi visto, optamos por escolher a Loja 04, pois além de ter um número grande de vendas, também está evoluindo no decorrer dos anos, mostrando potencial para expansão.

[Link do Colab:](https://colab.research.google.com/drive/1PapOJ3q16gNOZQlvNSV9DSIx34is_HZ7?usp=sharing)

https://colab.research.google.com/drive/1PapOJ3q16gNOZQlvNSV9DSIx34is_HZ7?usp=sharing



Estamos à disposição para qualquer dúvida!
Mikaely Pedrosa

