

Mini Projeto - Walmart

Deliverable 02 - Slot 04

RID: 5362 - Pollyane Rodrigues Fernandes - **Product Owner**

RID: 7576 - Rafael Dos Santos Geonmonond - **Scrum Master**

RID: 1587 - Danyelle De Barros Benatti - **Dev Team**

RID: 4692 - Marcelo Reis dos Santos - **Dev Team**

RID: 13302 - Lucas Alexandre Barbosa De Oliveira Santos - **Dev Team**

RID: 7645 - Letícia Campos Valente - **Dev Team**

RID: 7221 - Tamara Dos Anjos - **Dev Team**

Contextualização

Você foi contratado pelo **Wallmart** para fazer um levantamento do faturamento das lojas nos USA e apontar qual loja seria melhor para **expandir seu tamanho**.

É preciso analisar as vendas semanais de cada loja, calcular algumas informações importantes que serão perguntadas e ao final de tudo, **indicar em qual loja deva ser investida**.

Walmart Inc.

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.

O Que Devo Fazer?

A empresa te contratou para fazer um levantamento do faturamento das lojas nos USA e apontar qual loja seria melhor para **expandir seu tamanho**.

Você deve montar um colab bem redigido mostrando o passo a passo das etapas de análise do case, fazendo uma introdução e dados atuais sobre o problema.

Também é importante ressaltar/explicar passagens complexas ou insights que você tiver durante a resolução desse projeto.

Sobre os Dados

Dataset: <https://www.kaggle.com/yasserh/walmart-dataset>

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 descrição destas características estão abaixo:

Holiday Events	Semanas
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

Variável	Descrição
Store	Número da Loja
Date	Semana de Venda
Weekly_Sales	Venda Naquela Semana
Holiday_Flag	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

Dicas de Insights

1. Qual o preço médio de venda semanal por loja?
1. Qual o atributo que tem maior correlação (positiva ou negativa) com o número de vendas? (E que faça sentido)
1. Da loja com maior venda acumulada (soma de vendas de todo o período), quantas semanas do ano ele ultrapassou a média do período?
1. Faça um gráfico com as vendas máximas, mínimas e médias de todas as lojas ao longo do tempo. Ressalte de alguma forma, o período de feriados mencionados acima.
1. Pensando na expansão de uma das lojas, que loja você escolheria e o por que?

Obs: Além desses insights exploramos outros gerados pelo próprio grupo.

A Questão de Negócio:

META

A melhor loja Walmart dos EUA para **expandir o tamanho**.



Conhecimento de
negócio*



Limpeza e Análise exploratória
dos dados



Conclusões a partir da análise
exploratória

CHEGAREMOS NA META UNINDO AS
TRÊS FRENTES

*Perguntas fornecidas pela DNC
(Dicas de insights)

O Entendimento do Negócio:

2.1 Quais dados eu tenho disponível?

No banco de dados temos 45 lojas da Walmart dos USA, esta empresa é composta por um período de aproximadamente 3 anos (05/02/2010 a 26/10/2012), que estão agrupados por semanas, contendo:

- Datas dos maiores feriados;
- Número de lojas;
- Quais semanas houveram vendas e a quantidade;
- Temperatura da região;
- Preço do combustível da região onde está a loja;
- Índice do Preço do Consumidor (inflação);
- Taxa de desemprego.

O Entendimento do Negócio:

2.2 Eles são relevantes para o problema?

Problema: (1-Levantamento do faturamento das lojas e 2 Qual é a melhor loja para se investir).

1- Valor das vendas por semana nos diz o faturamento.

2- Taxa de desemprego é um ponto interessante para a análise, para sabermos se naquela região há baixo desemprego, se a economia do local está em alta e se faz sentido investir para expandir o negócio. Analisar se é uma região onde já possui grande número de lojas, também vale avaliar. Ver se têm correlação a temperatura com a quantidade de vendas.

2.3 Eles me trazem uma solução direta?

Do faturamento sim, mas de qual é a melhor empresa para expandir não, os dados por si só não trazem uma resposta. É preciso cruzar os dados para ter algumas interpretações e fazer comparação entre os períodos.

2.4 Qual o meu setor ou range de negócio?

O setor é multinacional, varejo em geral que engloba eletrodomésticos, setor alimentício, se ramificou em diversos novos mercados, tem suas próprias farmácias, serviços de automóveis, joalherias, etc.

A Coleta de Dados:

3.1 Os dados que eu tenho fazem sentido?

Sim. O conjunto disponibilizado reúne os dados de vendas semanais de 45 lojas americanas da Walmart, bem como registros de fatores que podem estar correlacionados com o faturamento. Além de fazerem sentido, os dados são de qualidade. Não foram observados, por exemplo, registros com valores nulos.

3.2 Estão no formato que eu gostaria?

Sim. A maioria das colunas estão no formato numérico, portanto podem ser utilizadas para descrições e cálculos iniciais. Apenas uma coluna (Date) foi transformada, seus valores estavam no formato 'objeto' e foram convertidos para o formato 'datetime64'.

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

Como os registros foram disponibilizados em relação ao faturamento semanal por loja, é possível agrupar os valores de cada loja para se ter um panorama geral das unidades. Com essa transformação é possível calcular, por exemplo, o faturamento total por loja, o faturamento médio semanal por loja, etc. Além disso, é possível identificar quais variáveis estão mais associadas ao faturamento, o que pode auxiliar na identificação de lojas promissoras para futuras expansões.

A Limpeza de Dados:

4.1 - Remoção de dados outliers e valores ausentes:

Aplicamos comandos como `isnull().sum()`, `uplicated` e `info()`, para entender se tínhamos dados Null, duplicados e se as colunas estavam com seus tipos de dados corretos.

Na análise não encontramos dados Nulos ou duplicados, mas identificamos que o tipo da coluna `Date` estava como `Object`, nesse caso convertermos para o tipo de dado correto (`Date`), utilizando o comando `.astype('datetime64[ns]')`.

```
1 df_duplicadas = df[df.duplicated(keep='first')]
2 print(df_duplicadas)
```

Empty DataFrame

Columns: [Date, Weekly_Sales, Holiday_Flag, Temperature, Fuel_Price, CPI, Unemployment]

Index: []

```
1 df.isnull().sum()
```

```
Date      0
Weekly_Sales  0
Holiday_Flag  0
Temperature  0
Fuel_Price  0
CPI         0
Unemployment  0
dtype: int64
```

```
1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 6435 entries, 14 to 33
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Date        6435 non-null   object
1   Weekly_Sales 6435 non-null   float64
2   Holiday_Flag 6435 non-null   int64
3   Temperature  6435 non-null   float64
4   Fuel_Price   6435 non-null   float64
5   CPI          6435 non-null   float64
6   Unemployment 6435 non-null   float64
dtypes: float64(5), int64(1), object(1)
memory usage: 402.2+ KB
```

A Exploração de Dados + Conclusão Final

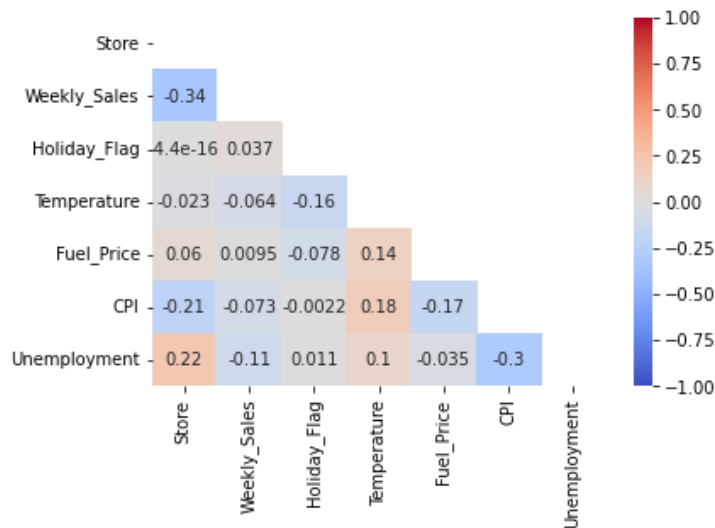
Utilizando os insights fornecidos pela DNC, conseguimos chegar a uma análise em que as quatro lojas que tiveram maior faturamento foram: 20, 4, 14 e 13. Considerando que os valores de faturamento médio e total das 4 melhores lojas são muito similares, então decidimos explorar as informações de desemprego e CPI para as quatro lojas (mínimo, médio e máximo).

Além disso, investigamos o faturamento destas 4 lojas ao longo do tempo para verificar a tendência de comportamento de vendas nestes locais.

Store	Weekly_Sales		CPI			Unemployment		
	Sum	Mean	Max	Mean	Min	Max	Mean	Min
20	3,01E+08	2,10E+06	216,15	209,04	203,61	8,18	7,37	6,96
4	2,99E+08	2,09E+06	131,19	128,67	126,06	8,62	5,96	3,87
14	2,89E+08	2,02E+06	192,33	186,29	181,64	8,99	8,65	8,42
13	2,86E+08	2,00E+06	131,19	128,68	126,06	8,30	7,00	5,62

A Exploração de Dados + Conclusão Final

A partir da correlação entre vendas semanais das lojas, tem-se que os fatores que mais afetam as vendas semanais são o desemprego e o índice de preços ao consumidor, com correlação de -0,11 e -0,073, respectivamente.

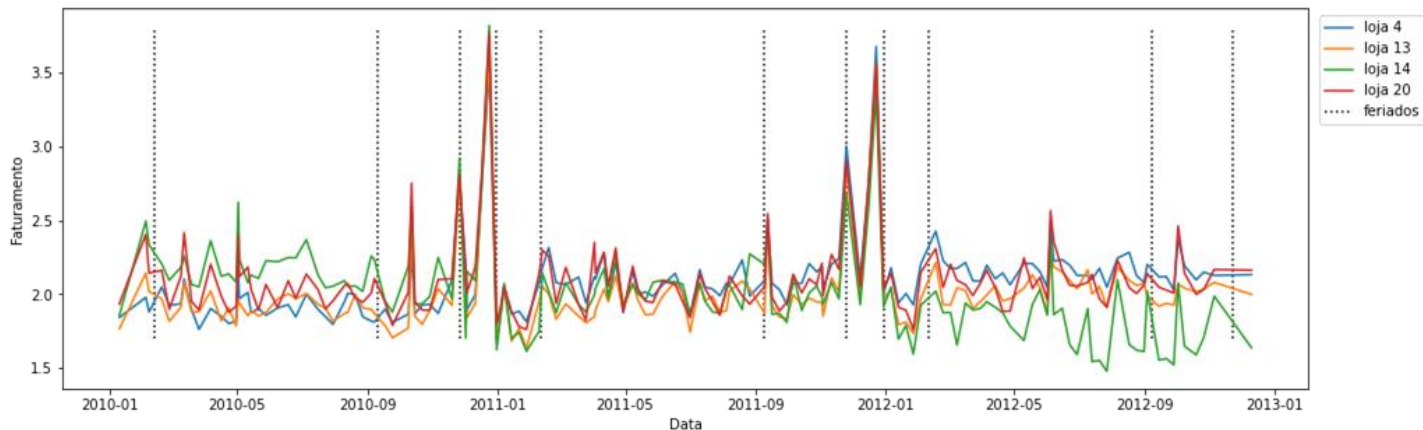


O CPI mede a variação de preços a partir da perspectiva do consumidor, sendo uma maneira fundamental para medir as variações de tendências de compra e inflação nos Estados Unidos.

A Exploração de Dados + Conclusão Final

Baseado nesta exploração, concluímos que a loja 4 seria a loja mais adequada para investimento, pois apresentou faturamento muito semelhante a loja 20 (2º lugar, o que é muito bom), porém tem menores índices de desemprego e índice de preço ao consumidor.

A partir do gráfico de faturamento, a loja 4 aparentemente vem aumentando seu faturamento ao longo do tempo (curva azul), gerando bons indicadores de crescimento. Diferentemente da loja 14, que apresenta claro comportamento de queda.



A Exploração de Dados + Conclusão Final

Corroborando com os dados anteriores, foi verificado que a loja 4 apresentou liderança com folga em máximo de faturamento semanal, das 143 semanas analisadas, ela liderou em 60 semanas (~42%), como pode ser evidenciado abaixo:

Loja	Nº de semanas em que liderou o faturamento
4	60
20	37
14	44
13	1

Neste sentido, escolhemos recomendar a loja 4 para investimento.

Perspectivas futuras

Depois das análises realizadas entendeu-se que os dados principais não eram suficientes em uma análise de expansão comercial, concluiu-se, então, que para análises **definitivas** outros fatores deveriam ser levados em consideração, como:

- Lucro;
- Tamanho da loja;
- Localização;
- Fluxo de pessoas;
- Custos etc.

Apesar disso, acredita-se veementemente nos resultados apresentados, portanto, para o cliente, indicáramos a **Loja 4**.



OBRIGADO!



©dnc group. Todos os direitos reservados.