



UNIVERSIDADE PRESBITERIANA MACKENZIE

## PROJETO APLICADO I – CURSO CIÊNCIA DE DADOS

TURMA 201825166.000.02 – GRUPO PROJETO APLICADO 3

GUILHERME AUGUSTO LEAL OLIVEIRA

GUILHERME ROCHA DE SOUZA DUARTE

GUILHERME SANTOS OLIVEIRA

GUSTAVO DA CONCEIÇÃO GUIMARÃES

RICARDO ZULIAN DE SOUZA AMARAL

## ANÁLISE EXPLORATÓRIA DE DADOS - WALMART

São Paulo

2025

TURMA 201825166.000.02 – GRUPO PROJETO APLICADO 3

GUILHERME AUGUSTO LEAL OLIVEIRA

GUILHERME ROCHA DE SOUZA DUARTE

GUILHERME SANTOS OLIVEIRA

GUSTAVO DA CONCEIÇÃO GUIMARÃES

RICARDO ZULIAN DE SOUZA MARAL

## ANÁLISE EXPLORATÓRIA DE DADOS - WALMART

Projeto aplicado apresentado à Universidade Presbiteriana Mackenzie como requisito parcial para conclusão da disciplina Projeto Aplicado I.

Orientador: Professor Lucas Cerqueira Figueiredo

São Paulo

2025

## 1 - SUMÁRIO

2 - TABELAS, QUADROS E FIGURAS.....	3
2.1 - QUADROS.....	3
3 - TERMOS CHAVE.....	3
4 - GLOSSÁRIO.....	3
5 - RECURSOS EXTERNOS .....	4
6 - INTRODUÇÃO.....	4
6.1 – A EMPRESA.....	4
7 – OBJETIVO .....	4
7.2 OBJETIVOS ESPECÍFICOS.....	5
8 - A BASE DE DADOS.....	5
9 - ANÁLISE EXPLORATÓRIA .....	6

## 2 - TABELAS, QUADROS E FIGURAS

### 2.1 - QUADROS

Quadro 1 – Campos do Dataset.....	5
Quadro 2 – Resumo da Base.....	5

## 3 - TERMOS CHAVE

Vendas, sazonalidade, fatores socioeconômicos.

## 4 - GLOSSÁRIO

**CPI** – Sigla para Customer Price Index, ou o índice de inflação acumulada na semana. É um número inteiro representando o valor da cesta de produtos medidos em relação a uma data base, que tem valor 100. Um CPI de 110, por exemplo, indica uma inflação de 10% no período.

**Holiday Flag** – Indica se a semana analisada contém um feriado.

**Weekly Sales** – Vendas semanais da loja em dólares americanos.

**Boom, Bust e Neutral** – Estrondo, falência e neutro, jargão inglês refletindo fases de ciclos de negócios.

## 5 - RECURSOS EXTERNOS

Os documentos e o código desenvolvidos para a realização deste estudo podem ser encontrados no Github.

Segue o repositório: <https://github.com/guilhermersduarte/Projeto-Aplicado-1>

## 6 - INTRODUÇÃO

Este projeto de análise exploratória de dados tem como foco o Walmart, uma das maiores redes varejistas do mundo. O objetivo é investigar padrões e tendências em dados relacionados às vendas das lojas da empresa e quais fatores afetam sua performance. Utilizando bases de dados públicas, serão analisadas variáveis como volume de vendas por loja, sazonalidade, influência de inflação, juros, desemprego - entre outros.

Ferramentas como Python e R serão empregadas para limpeza, visualização e interpretação dos dados. A análise busca responder perguntas como: quais fatores influenciam as vendas? Qual a velocidade de resposta das vendas às alterações nas condições socioeconômicas? Os resultados esperados incluem insights acionáveis para otimização de estoque, formação de preço e estratégias de marketing. O projeto também pode servir como base para estudos futuros envolvendo previsão de vendas.

### 6.1 – A EMPRESA

A história do Walmart tem início em 1950, quando Sam Walton comprou uma loja e a inaugurou como Walton's Five and Dime. A rede Walmart propriamente dita foi fundada em 1964 com a abertura de uma única loja em Rogers, Arkansas.

O Walmart tem como missão “ajudar as pessoas a economizarem dinheiro para que possam viver melhor”. Seus valores incluem integridade, respeito ao indivíduo e compromisso com os clientes.

O Walmart é uma gigante do setor varejista, com 2,1 milhões de funcionários e 10.771 lojas ao redor do mundo (2025).

O Walmart emprega ferramentas de análise de dados para prever demandas, otimizar estoques e personalizar ofertas.

## 7 – OBJETIVO

O estudo visa analisar e explorar os dados de vendas semanais e de fatores que podem afetar o desempenho das lojas do Walmart, identificando padrões em vendas, sazonalidade e impactos de variáveis socioeconômicas como inflação, desemprego, preço de combustível e outros fatores buscando oferecer insights estratégicos

Para isso estudaremos o comportamento das vendas nas duas dimensões oferecidas: No

tempo e por loja individual. Depois cada variável individual será estudada para avaliar a sua influência no volume de vendas.

Por último serão oferecidas soluções para a administração de pessoal e estoque das lojas para atender essas variações de vendas bem como criar ações que possam mitigar (em caso de queda) ou potencializar (em caso de aumento) o efeito das variáveis sobre as vendas.

## **7.2 OBJETIVOS ESPECÍFICOS**

O Estudo se dividirá em quatro etapas com objetivos e entregas definidas:

### **1. Preparação:**

Na primeira etapa o grupo inicia a criação do repositório na plataforma GitHub, organizando a base do projeto. É nessa fase que ocorre a definição da empresa e o contexto da análise, garantindo que fique claro o propósito do trabalho. Também é feita uma breve análise preliminar de todos os objetivos, colunas, descrição do dataset e a criação do calendário.

### **2. Análise exploratória e desenvolvimento de propostas:**

Na segunda entrega, o foco é na análise exploratória de dados, que inclui a avaliação das vendas semanais e identificação de correlações como temperatura, preços do combustível, CPI e desemprego. essa análise tem como objetivo compreender como esses fatores impactam o desempenho das lojas e se há padrões, trazendo uma proposta analítica mais concisa e completa.

### **3. Storytelling e comunicação dos resultados:**

Na terceira entrega temos como objetivo trabalhar com o storytelling dos dados apresentados, desenvolvendo narrativas dos insights desenvolvidos na segunda fase. Para isso, serão revisaremos os scripts e estruturas que desenvolvemos, elaborando uma estratégia visual para apresentação dos resultados, criando um dashboard que destaque as tendências e padrões.

### **4. Conclusão e apresentação:**

Nesta última etapa, o grupo apresentará um vídeo com a narrativa dos dados juntamente com o relatório final, incluindo todas as conclusões analíticas e estratégias.

## **8 - A BASE DE DADOS**

Selecionamos uma base pública no Kaggle chamada Walmart Sales, publicada por Mikhail. A base engloba 6435 registros de vendas semanais em 45 lojas do Walmart num período de 143 semanas.

Os dados foram coletados entre 05/02/2010 e 26//10/2012. Entendemos que apesar da idade considerável da amostra é válido estudá-la, uma vez que buscamos entender a reação

das vendas à variação de dados ambientais e socioeconômicos, e não a relação das vendas com números absolutos que ficaram obsoletos.

Uma análise preliminar em R mostra dados coesos, sem nulos. Fica patente a necessidade de conversão do formato de data no campo 'DATE', que a importação em R não entendeu como datas.

Não existem dados sensíveis, tais como nomes e atributos de identificação de pessoas ou unidades de negócios.

A base em sua forma original contém 8 colunas, como descrito no Quadro 1, a seguir:

Quadro 1 – Campos do dataset Walmart Sales		
Nome da Coluna	Tipo de Dado	Descrição.
STORE	Numérico	Referência ao número da loja representada na linha.
DATE	Texto	Texto representando o dia em que se inicia a semana representada na linha, no formato dd-mm-yyyy.
WEEKLY_SALES	Numérico	Apresenta o total de vendas semanal em dólares americanos.
HOLIDAY_FLAG	Binário	indica a ocorrência de feriado na semana representada na linha.
TEMPERATURE	Numérico	Representa a temperatura média em graus fahrenheit na semana.
FUEL_PRICE	Numérico	Indica o preço médio do combustível na região- em dólares por galão.
CPI	Numérico	Indica a inflação acumulada no período
UNEMPLOYMENT	Numérico	Representa o desemprego na semana, na região em pontos percentuais com uma casa decimal
Fonte: Elaborado pelos autores.		

## 9 - ANÁLISE EXPLORATÓRIA

Inicialmente, verificamos o resumo estatístico da base de dados buscando saber o número e valores não nulos, que foi de 6.435 para todas as variáveis, também descobrimos os valores Mínimos, Médios, Máximos a Variância e o Desvio padrão de cada coluna, resumido na seguinte tabela:

Quadro 2 – Resumo estatístico da base de dados					
	Mínimo	Média	Máximo	Variância	Desv Padrão
Weekly_Sales (US\$mm)	0.2099	1.0469	3.8186	0.3185	0.5643
Temperature (°C)	-18.9222	15.9243	37.8555	105.0047	10.2471
Fuel_Price (U\$)	2.4720	3.3586	4.4680	0.2106	0.4590
CPI (%)	126.0640	171.5783	227.2328	1548.9508	39.3567
Unemployment (%)	3.8790	7.9991	14.3130	3.5189	1.8758
Fonte: Elaborado pelos autores.					

\*Dividimos o valor da coluna Weekly\_Sales por 1.000.000 para facilitar a visualização.

Quando verificamos a distribuição das colunas, conseguimos apontar que a maior parte das vendas semanais estão em torno de US\$ 500.000 e que faturamentos acima dos US\$ 2.500.000 são eventos raros. Na temperatura, é possível verificar que na maior parte das semanas, a temperatura ficou um pouco acima de 20 graus celsius. Assim como podemos concluir que o preço do combustível ficou por mais tempo na faixa de valor entre US\$ 3,50 e US\$ 3,75.

Através da análise exploratória da base de dados de vendas do Walmart, validou a sua integridade, verificando a ausência de valores nulos que impactem na análise ou inconsistências nos dados das principais variáveis. A conversão do campo de data foi necessária para viabilizar análises temporais, assim como a conversão do campo de Temperatura de fahrenheit para graus celsius. Com um total de 45 lojas e aproximadamente 143 semanas, a base permite observar o impacto causado nas vendas semanais por 5 variáveis, temperatura, preço do combustível taxa de juros, desemprego e feriados. Inicialmente, definimos 3 para explorarmos: Feriados, Desemprego e Temperatura

## 9.2 Impacto dos Feriados

O impacto dos feriados nas vendas foi um dos primeiros pontos a ser investigado, utilizando a variável [Holiday Flag] que indica se teve algum feriado relevante naquela semana. A média geral de vendas em semanas sem feriados foi de aproximadamente US\$ 1.041.256, enquanto a média geral de vendas em semanas com feriados foi de aproximadamente US\$ 1.122.888, indicando um aumento de 7,8% nas vendas em semanas com feriados.

Aprofundando a análise, criamos uma segmentação adicional das semanas em torno dos feriados buscando criar uma análise que consiga identificar oportunidades estratégicas específicas para cada unidade, classificando como: **Boom:** semanas nas **3 semanas anteriores** a um feriado, **Bust:** semanas nas **3 semanas posteriores** a um feriado, **Neutral:** semanas fora dessas janelas.

Aplicando essa classificação é possível identificar padrões comportamentais visualizando que em boa parte das lojas, semanas Boom apresentam vendas superiores a semanas classificadas como Neutral demonstrando um padrão de antecipação nas compras, assim como demonstra uma queda significativa em semanas Bust, com as vendas ficando abaixo da média. O efeito Boom e Bust varia conforme a loja, demonstrando também que o comportamento do consumidor tem características regionais.