



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

SUMÁRIO

INTRODUÇÃO	4
TERMOS CHAVE	4
GLOSSÁRIO	4
RECURSOS EXTERNOS.....	4
BASE DE DADOS	5

LISTA DE QUADROS

Quadro 1 – Campos do Dataset.....	5
-----------------------------------	---

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.

TERMOS CHAVE

Vendas, sazonalidade, fatores socioeconômicos.

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.

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>

A BASE DE DADOS

Selecionamos uma base pública no Kaggle chamada Walmart Sales, publicada por Mikhail. A base engloba as vendas semanais em 45 lojas do Walmart num período de 148 semanas com início de 05/02/2010 a 26//10/2012.

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: Elaboradp pelos autores.		

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.

Fica aberta a possibilidade de acrescentarmos uma coluna relativa a taxa básica de juros a partir de outra fonte de dados.

REFERÊNCIAS

Mikhail - KAGGLE. *Walmart sales*. Disponível em:

<https://www.kaggle.com/datasets/mikhail1681/walmart-sales?resource=download>.

Acesso em: 24 fev. 2025.