Logotipo, Ícone

O conteúdo gerado por IA pode estar incorreto.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](#_Toc193788241)

[2.1 - QUADROS 3](#_Toc193788242)

[3 - TERMOS CHAVE 3](#_Toc193788243)

[4 - GLOSSÁRIO 3](#_Toc193788244)

[5 - RECURSOS EXTERNOS 3](#_Toc193788245)

[6 - INTRODUÇÃO 4](#_Toc193788246)

[6.1 – A EMPRESA 4](#_Toc193788247)

[7 – OBJETIVO 4](#_Toc193788248)

[7.2 OBJETIVOS ESPECÍFICOS 5](#_Toc193788249)

[8 - A BASE DE DADOS 5](#_Toc193788250)

[9 - ANÁLISE EXPLORATÓRIA 6](#_Toc193788251)

# 2 - TABELAS, QUADROS E FIGURAS

## 2.1 - QUADROS

Quadro 1 – Campos do Dataset 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.

# 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 variáveissocioeconômicas 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.

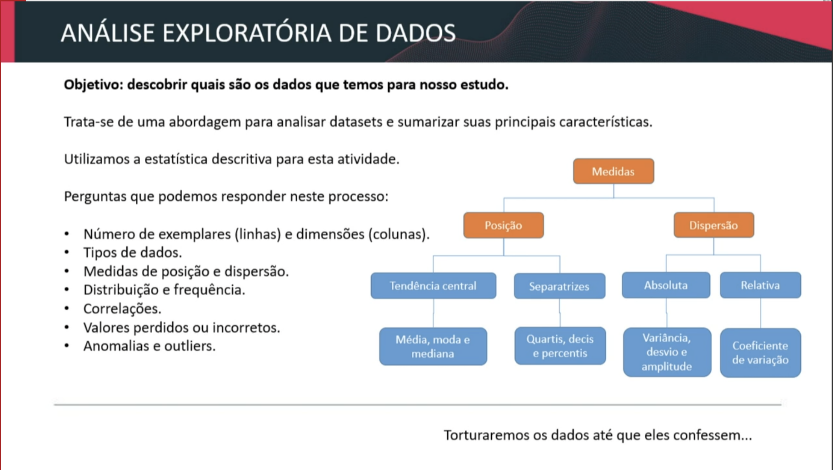
Uma análise preliminar em R mostra dados coesos, sem nulos. Fica patente a necessidade de conversão do formato de data no campo ‘DATE’, que a importação em R não entendeu como datas.

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



|  |  |
| --- | --- |
| rganização do Documento | **1.25 pts** |
| **ÓTIMO** |
| Documento possui capa, sumário, lista de figuras, lista de tabelas, introdução e capítulos separados para cada conjunto de itens solicitados. |
| **0.75 pts** |
| **BOM** |
| Documento possui a maior parte dos itens solicitados. |
| **0.5 pts** |
| **REGULAR** |
| Documento possui somente alguns dos itens solicitados. |
| **0 pts** |
| **INSATISFATÓRIO** |
| Documento está desorganizado e sem nenhuma estrutura identificável. |
| Seção de Proposta Analítica inserida no Documento | **1.25 pts** |
| **ÓTIMO** |
| A seção foi inserida e apresenta os itens conforme apresentado na videoaula, nos slides sobre apresentação da Organização/Empresa, Problema do Estudo, Metadados e a Proposta do que será desenvolvido e Pipeline de dados para apoiar o processo analítico. |
| **0.75 pts** |
| **BOM** |
| A seção foi inserida, mas faltam alguns itens para atender a totalidade dos itens solicitados. |
| **0.5 pts** |
| **REGULAR** |
| A seção foi inserida, mas faltam vários itens para atender a totalidade dos itens solicitados. |
| **0 pts** |
| **INSATISFATÓRIO** |
| A seção foi inserida, mas precisa ser reformulada pois não atende ao que foi solicitado. |
| Scripts da Análise Exploratória em Python | **1.25 pts** |
| **ÓTIMO** |
| Os scripts foram inseridos no GitHub e atendem ao que foi solicitado para compor a Análise Exploratória dos Dados do Projeto. Os scripts são notebooks Jupyter, escritos em Python e foram desenvolvidos com comandos, bibliotecas e linhas de comentários, atendendo a boas práticas de organização do código. |
| **0.75 pts** |
| **BOM** |
| Os scripts foram inseridos no GitHub e atendem ao que foi solicitado para compor a Análise Exploratória dos Dados do Projeto. Os scripts apresentam algumas falhas em documentação ou na execução dos comandos. |
| **0.5 pts** |
| **REGULAR** |
| Os scripts foram inseridos no GitHub e atendem ao que foi solicitado para compor a Análise Exploratória dos Dados do Projeto. Os scripts apresentam várias falhas em documentação ou na execução dos comandos. |
| **0 pts** |
| **INSATISFATÓRIO** |
| Os scripts foram inseridos no GitHub mas apresentam várias inconformidades. |
| Seção de Análise Exploratória inserida no Documento | **1.25 pts** |
| **ÓTIMO** |
| A seção foi inserida e apresenta os itens alinhados aos scripts que foram inseridos no GitHub. O texto apresenta uma descrição das variáveis, complementando o metadados que já havia sido apresentado, para cada coluna foram descritos itens como número de exemplares, valor máximo/mínimo, variância, desvio, distribuição, quantidade de NAs, existência de outlayers, entre outros. Utilizou-se alguns gráficos para melhor detalhar a amostra. |
| **0.75 pts** |
| **BOM** |
| A seção foi inserida, mas faltam alguns itens para atender a totalidade dos itens solicitados. |
| **0.5 pts** |
| **REGULAR** |
| A seção foi inserida, mas faltam vários itens para atender a totalidade dos itens solicitados. |
| **0 pts** |
| **INSATISFATÓRIO** |
| A seção foi inserida, mas precisa ser reformulada pois não atende ao que foi solicitado. |

10 - REFERÊNCIAS

Mikhail - KAGGLE. *Walmart sales*. Disponível em: <https://www.kaggle.com/datasets/mikhail1681/walmart-sales?resource=download>. Acesso em: 24 fev. 2025.