

UNIVERSIDADE ESTÁCIO DE SÁ
CAMPUS NOVA AMÉRICA

TÓPICOS DE BIG DATA EM PYTHON

Victor Costa de Oliveira

Orientador(a): Lucas Floriano

2025

Campus Nova América – RJ, Rio de Janeiro

Sumário

1. DIAGNÓSTICO E TEORIZAÇÃO	3
1.1. Identificação das partes interessadas e parceiros	3
1.2. Problemática e/ou problemas identificados	6
1.3. Justificativa	6
1.4. Objetivos/resultados/efeitos a serem alcançados (em relação ao problema identificado e sob a perspectiva dos públicos envolvidos)	6
1.5. Referencial teórico (subsídio teórico para propositura de ações da extensão)	7
2. PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO	8
2.1. Plano de trabalho (usando ferramenta acordada com o docente)	8
2.2. Descrição da forma de envolvimento do público participante na formulação do projeto, seu desenvolvimento e avaliação, bem como as estratégias pelo grupo para mobilizá-los.	10
2.3. Grupo de trabalho (descrição da responsabilidade de cada membro)	11
2.4. Metas, critérios ou indicadores de avaliação do projeto	11
2.5. Recursos previstos	12
2.6. Detalhamento técnico do projeto.....	12
3. ENCERRAMENTO DO PROJETO	14
3.1. Relatório Coletivo (podendo ser oral e escrita ou apenas escrita).....	14
3.2. Avaliação de reação da parte interessada	14
3.3. Relato de Experiência Individual	17
3.1. CONTEXTUALIZAÇÃO	17
3.2. METODOLOGIA	18
3.3. RESULTADOS E DISCUSSÃO:	19
3.4. REFLEXÃO APROFUNDADA	23
3.5. CONSIDERAÇÕES FINAIS.....	24

1. DIAGNÓSTICO E TEORIZAÇÃO

1.1. Identificação das partes interessadas e parceiros

O projeto realizado terá como parte interessada o dono da empresa MateGuaraná, uma loja voltada ao comércio de doces e salgados, é composta por um público diversificado, abrangendo pessoas de todos os gêneros e faixas etárias. A empresa apresentou uma necessidade que logo foi acordada entre ambas as partes.



CARTA DE APRESENTAÇÃO

Vimos por esta apresentar o grupo de acadêmicos da Universidade Estácio de Sá – UNESA listado na tabela a final deste documento, a fim de convidá-lo a participar de uma atividade extensionista associada à disciplina TÓPICOS DE BIG DATA EM PYTHON sob responsabilidade do Prof. Lucas Antunes Floriano.

Em consonância ao Plano Nacional de Educação e demais normativas educacionais vigentes, a Universidade Estácio de Sá – UNESA desenvolve atividade extensionista que, norteados pela metodologia de aprendizagem baseada em projetos, tem por princípios fundantes o diagnóstico dos problemas/demandas/necessidades, a participação ativa dos interessados/públicos participantes, a construção dialógica, coletiva e experiencial de conhecimentos, o planejamento de ações, o desenvolvimento e avaliação das ações, a sistematização dos conhecimentos, a avaliação das ações desenvolvidas.

Nesse contexto, a disciplina acima mencionada tem como principal escopo os temas relacionados à BIG DATA, com foco em aplicações em python.

Sendo assim, pedimos o apoio dessa organização/entidade/coletivo/associação/outro, que aqui chamaremos de parte interessada, para a realização das seguintes atividades: diagnósticos, análises, entrevistas, levantamentos, projetos ou qualquer outra metodologia de estudo de caso que auxilie no desenvolvimento das competências de nossos acadêmicos e ao mesmo tempo possa contribuir para a comunidade em que estamos inseridos.

Como se trata de atividade de ensino/aprendizagem de caráter extensionista, prevista no Projeto Pedagógico do Curso, salientamos que:

- não há cobrança de remuneração de qualquer natureza por parte da Universidade Estácio de Sá, seus alunos ou o docente da disciplina, à parte interessada;
- as atividades desenvolvidas no âmbito do projeto extensionista não configuram relação de trabalho entre os alunos e o docente da Universidade Estácio de Sá – UNESA disciplina TÓPICOS DE BIG DATA EM PYTHON, e a parte interessada;
- os resultados do projeto só poderão ser implantados para uso efetivo mediante Anotação de Responsabilidade Técnica de um profissional habilitado;
- os resultados do projeto podem ser implantados pela parte interessada para fins lucrativos, sem a necessidade de pagamento de quaisquer benefícios aos alunos, ao docente da disciplina e à Universidade Estácio de Sá – UNESA;
- quaisquer custos relativos à implantação e operação contínua do projeto fora do escopo das atividades do presente projeto serão arcados pela parte interessada;
- Poderá ser solicitado à parte interessada o compartilhamento de dados não-sensíveis a serem usados durante o desenvolvimento do projeto em BIG DATA, e nunca fora deste escopo.

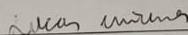
- Poderá ser solicitado à parte interessada o compartilhamento de dados não-sensíveis a serem usados durante o desenvolvimento do projeto em BIG DATA, e nunca fora deste escopo.

Aproveitamos a oportunidade e solicitamos que, em caso de aceite, seja formalizado, mediante assinatura da Carta de Autorização, as atividades e informações que o(s) aluno(s) poderá(ão) ter acesso.

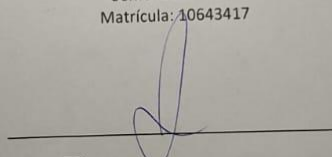
Desde já nos colocamos à sua disposição para quaisquer esclarecimentos. Professor Lucas Antunes Floriano- 98820-4968 e/ou luкас.floriano@estacio.br e -----

Grupo de Alunos
<Victor Costa de Oliveira> Matrícula: < 202402369989 >
<NOME COMPLETO DO ALUNO 2> Matrícula: <MATRÍCULA DO ALUNO 2>
<NOME COMPLETO DO ALUNO 3> Matrícula: <MATRÍCULA DO ALUNO 3>
<NOME COMPLETO DO ALUNO 4> Matrícula: <MATRÍCULA DO ALUNO 4>
<NOME COMPLETO DO ALUNO 5> Matrícula: <MATRÍCULA DO ALUNO 5>

Atenciosamente,



Lucas Antunes Floriano
Docente da disciplina: TÓPICOS DE BIG DATA EM PYTHON
Semestre: 2025.1
Matrícula: 10643417



<<parte interessada>>

Rio de Janeiro, 17 de Maio de 2025.

1.2. Problemática e/ou problemas identificados

Durante um debate com a parte interessada foi notado que a ausência de uma metodologia eficiente de tratamento e análise de dados impacta diretamente na produtividade e na tomada de decisões estratégicas para a empresa MateGuaraná, com a necessidade de realizar a contabilidade e armazenamento dos dados observados e informados de forma organizada e prática. A partir dessa demanda, o projeto foi desenvolvido com o objetivo de propor uma solução que permita realizar o tratamento adequado dos dados, gerando informações precisas, acessíveis e que contribuam diretamente para uma gestão mais eficiente e assertiva.

1.3. Justificativa

Este projeto tem grande importância para as disciplinas do curso, pois permite que os conhecimentos adquiridos em sala de aula sejam aplicados em um cenário real. Envolve diretamente o uso de ferramentas e linguagens muito aplicadas no mercado de tecnologia. Através do uso do Python e da biblioteca Pandas, foi possível realizar o tratamento, limpeza e organização de dados, atividades essenciais na área e extremamente presentes no mercado atual.

Além disso, a utilização de Power BI contribuiu para transformar esses dados em gráficos, relatórios e painéis interativos, facilitando a visualização e interpretação das informações. O trabalho com planilhas também se mostrou relevante, já que é uma das formas mais comuns de armazenamento de dados em pequenas e médias empresas. Com isso, o projeto proporciona uma experiência prática de como a programação é aplicada no contexto de análise e tratamento de dados

1.4. Objetivos/resultados/efeitos a serem alcançados (em relação ao problema identificado e sob a perspectiva dos públicos envolvidos)

Organizar e tratar os dados da empresa MateGuaraná, utilizando Python e Pandas, para facilitar a análise de informações como vendas, produtos e desempenho. Desenvolver relatórios e dashboards no Power BI que permitam uma visualização clara e eficiente dos dados, auxiliando na tomada de decisões. Automatizar processos de manipulação de dados, reduzindo erros manuais e otimizando o tempo gasto na organização das informações da empresa.

1.5. Referencial teórico (subsídio teórico para propositura de ações da extensão)

1. Este artigo “Big data: A review” ensina que Big data é um termo para conjuntos massivos de dados com estruturas grandes, mais variadas e complexas, com dificuldades de armazenamento, análise e visualização para processos ou resultados posteriores. O processo de pesquisa em grandes quantidades de dados para revelar padrões ocultos e correlações secretas denominadas análises de big data. Essas informações são úteis para empresas ou organizações com a ajuda de obter insights mais ricos e profundos e obter vantagem sobre a concorrência. Por esta razão, as implementações de big data precisam ser analisadas e executadas com a maior precisão possível. Este artigo apresenta uma visão geral do conteúdo, escopo, amostras, métodos, vantagens e desafios do big data e discute a preocupação com a privacidade nele

2. Neste artigo “Pandas: a foundational Python library for data analysis and statistics” é discutido pandas, uma biblioteca Python de recursos ricos estruturas de dados e ferramentas para trabalhar com conjuntos de dados estruturados comuns a estatísticas, finanças, ciências sociais e muitos outros campos. A biblioteca fornece rotinas integradas e intuitivas para realizar manipulações de dados comuns e análise de tais conjuntos de dados. O texto descreve que o Pandas foi desenvolvido com o objetivo de preencher uma lacuna existente na linguagem Python, que até então não possuía uma biblioteca dedicada e eficiente para tratamento e análise de dados tabulares e temporais. A biblioteca oferece estruturas de dados poderosas, como o DataFrame e a Series, que permitem trabalhar com dados de forma muito semelhante a softwares como Excel, SQL e até linguagens estatísticas como R.

Além disso, o artigo enfatiza que o Pandas fornece uma vasta gama de funções para limpeza, transformação, filtragem, agregação e análise de dados, tornando o processo mais rápido, prático e menos suscetível a erros. O uso de comandos simples e intuitivos permite que tanto programadores quanto analistas e pesquisadores, mesmo com pouco conhecimento em linguagens de programação, possam realizar análises complexas de maneira eficiente.

3. Já o artigo “Research data analysis with power bi” de Vijay Krishnan (2017), aborda o uso da ferramenta Power BI como um recurso eficiente para auxiliar na análise de dados em projetos acadêmicos e científicos. No trabalho, o autor demonstra como o Power BI pode ser aplicado para transformar grandes volumes de dados brutos em

informações visuais, por meio de gráficos, dashboards interativos e relatórios dinâmicos.

A publicação destaca as principais funcionalidades da ferramenta, como a importação de dados provenientes de diferentes fontes (planilhas, bancos de dados, arquivos da web, entre outros) e o tratamento desses dados de forma organizada para facilitar a interpretação e a tomada de decisões. Além disso, o artigo enfatiza que o Power BI não se limita ao uso empresarial, sendo extremamente útil no meio acadêmico, auxiliando pesquisadores e estudantes na organização, análise e visualização de dados de maneira prática, eficiente e visualmente compreensível.

Por meio de exemplos e aplicações práticas, o autor reforça que o uso de ferramentas de Business Intelligence, como o Power BI, contribui significativamente para melhorar a análise de dados em pesquisas, tornando os resultados mais claros, objetivos e acessíveis.

1. SAGIROGLU, Seref; SINANC, Duygu. Big data: A review. In: **2013 international conference on collaboration technologies and systems (CTS)**. IEEE, 2013. p. 42-47.
2. MCKINNEY, Wes et al. pandas: a foundational Python library for data analysis and statistics. **Python for high performance and scientific computing**, v. 14, n. 9, p. 1-9, 2011.
3. KRISHNAN, Vijay. Research data analysis with power bi. 2017.

2. PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO

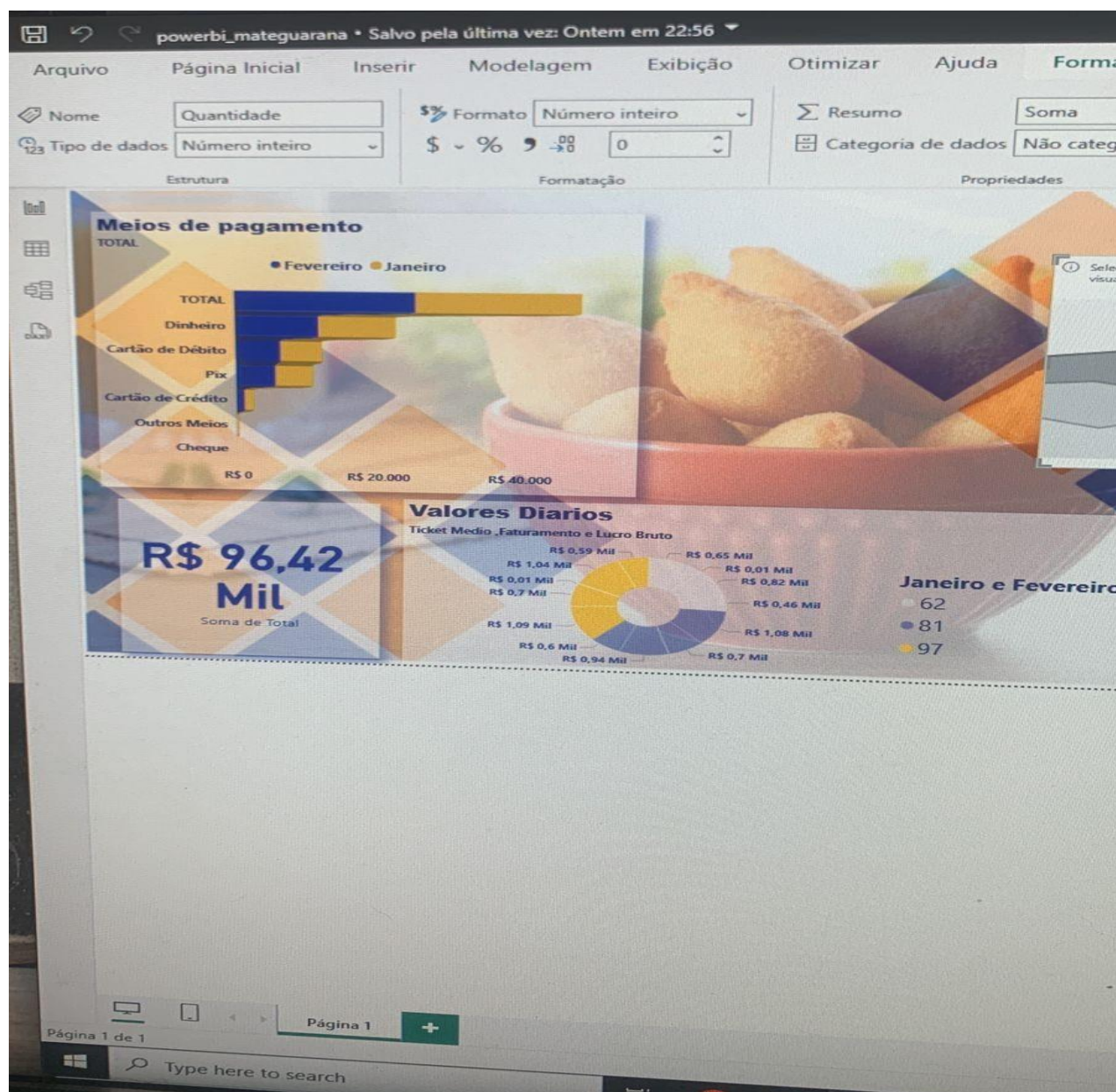
2.1. Plano de trabalho (usando ferramenta acordada com o docente)

Durante o desenvolvimento do projeto, os dados irão ser recebidos em formato estruturado, enviados pelo cliente principalmente através de planilhas Excel. Essas planilhas continham informações relacionadas a vendas, pedidos, controle de produção e dados financeiros da empresa.

Esses dados irão ser cuidadosamente analisados, tratados e organizados utilizando a linguagem de programação Python juntamente com a biblioteca Pandas, que permitiu realizar a limpeza, estruturação e transformação dos dados de forma eficiente.

Após o tratamento, os dados serão integrados em dashboards no Power BI, facilitando a visualização de informações essenciais para a tomada de decisões dentro da empresa. As análises produzidas ajudam a acompanhar o desempenho das vendas, identificar os produtos mais vendidos e fornecer uma visão clara do faturamento e despesas.

Foi realizado o tratamento de dados da loja e transformação em dashboard com power BI pelos integrantes do grupo:

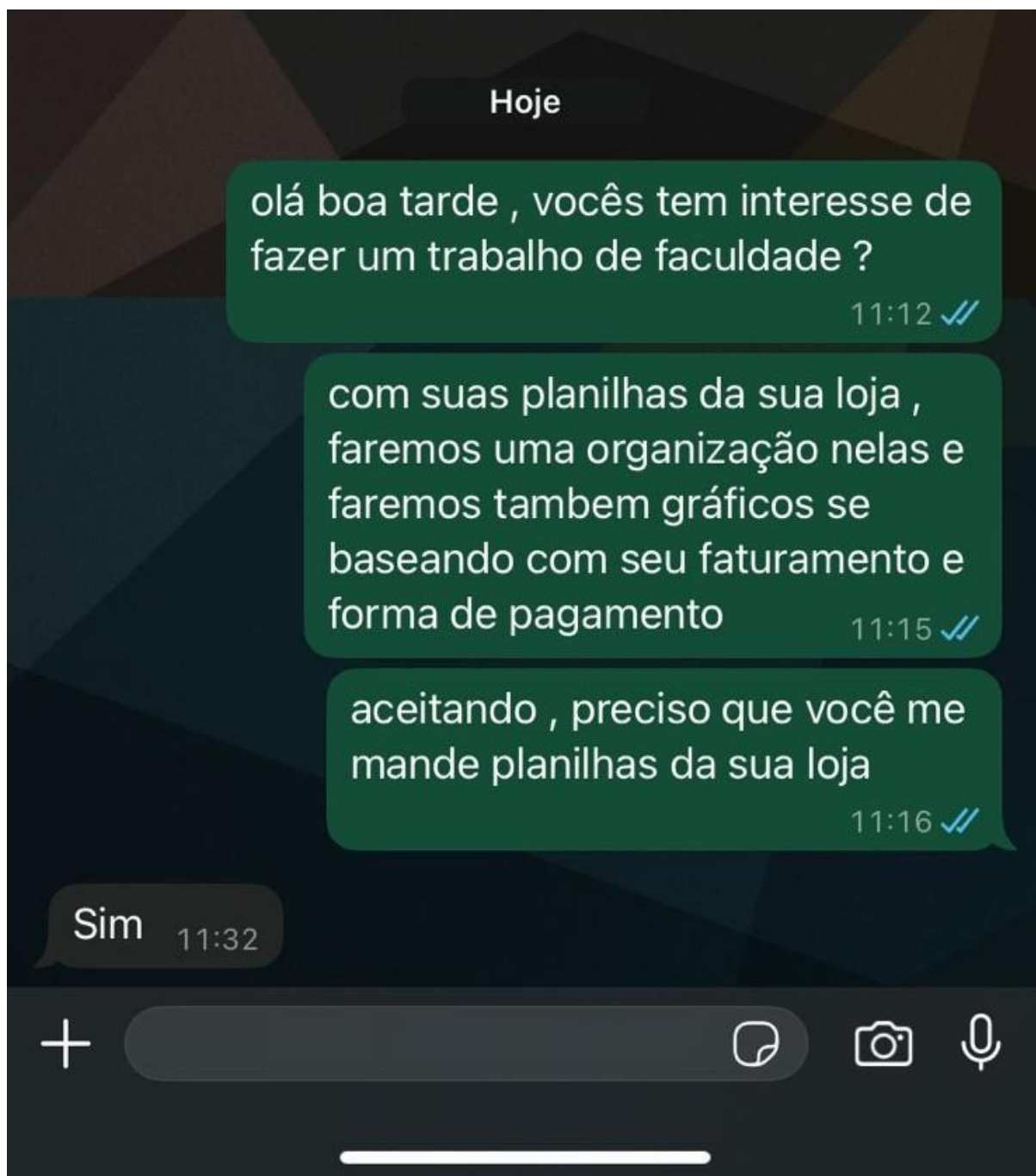


- **25/02** – Início do contato com o cliente (MateGuaraná) para alinhamento inicial
- **10/03** – Elaboração da carta de autorização e consentimento do trabalho
- **19/03** – Entrega da carta de autorização ao docente
- **25/03** – Organização dos dados fornecidos pelo cliente em planilhas e definição das métricas a serem analisadas
- **01/04** – Início do desenvolvimento dos scripts em Python utilizando Pandas para tratamento, limpeza e organização dos dados
- **25/04** – Desenvolvimento dos primeiros dashboards no Power BI com base nos dados tratados
- **15/05** – Ajustes finais nos scripts em Python, nos dashboards e validação final dos dados tratados
- **22/05** – Desenvolvimento dos textos dos itens 1 ao 3.1.1 do roteiro de extensão

- **30/05** – Preparação da apresentação, organização dos arquivos e finalização da documentação do projeto

2.2. Descrição da forma de envolvimento do público participante na formulação do projeto, seu desenvolvimento e avaliação, bem como as estratégias pelo grupo para mobilizá-los.

A comunicação com a empresa Mate guarana Doces e Salgados foi realizada de forma direta por meio de troca de mensagens, garantindo um contato constante durante todo o desenvolvimento do projeto. Esse diálogo facilitou a compreensão das principais demandas da empresa, possibilitando que suas necessidades fossem ouvidas e consideradas em cada etapa do trabalho. Além disso, a empresa colaborou ativamente, fornecendo os dados necessários para análise, acompanhando o andamento das atividades e contribuindo com feedbacks importantes para ajustes e melhorias no projeto.



2.3. Grupo de trabalho (descrição da responsabilidade de cada membro)

Victor Costa de Oliveira – Responsável por todo o projeto, desde a comunicação com a empresa, até o desenvolvimento dos códigos e PowerBI.

2.4. Metas, critérios ou indicadores de avaliação do projeto

O objetivo é organizar e tratar os dados da empresa MateGuaraná, otimizando a gestão de informações e facilitando a análise de dados. As etapas foram definidas da seguinte forma:

Coleta e organização dos dados: Inicialmente, foram recebidos os dados da empresa em formatos como Excel e imagens. Esses dados foram analisados, organizados e estruturados para garantir sua correta utilização no sistema.

Tratamento e manipulação dos dados: Utilizando Python e a biblioteca Pandas, realizei a limpeza, padronização e organização dos dados. Essa etapa foi essencial para garantir a precisão das informações e reduzir erros, preparando-as para serem utilizadas nas análises e relatórios.

Desenvolvimento de dashboards e relatórios: Através do Power BI, foram criados painéis interativos e relatórios visuais que facilitam a compreensão dos dados de vendas, estoque e desempenho da empresa.

Automatização dos processos: Foram desenvolvidos scripts em Python para automatizar tarefas rotineiras de organização e atualização dos dados, minimizando erros manuais e otimizando o tempo do responsável pela empresa.

2.5. Recursos previstos

O projeto tem caráter totalmente educacional, sem fins lucrativos, sendo desenvolvido com o objetivo de aplicar na prática os conhecimentos adquiridos em sala de aula. Foram utilizados recursos gratuitos e acessíveis, como Python, Pandas, Power BI e planilhas, que auxiliaram na organização, tratamento e análise dos dados. Todo o desenvolvimento foi realizado de forma colaborativa, visando o aprendizado e o desenvolvimento de habilidades na área de programação e análise de dados. Foram utilizados recursos gratuitos, como:

- **Python:** Linguagem de programação utilizada para processar, limpar e organizar os dados.
- **Pandas:** Biblioteca do Python aplicada na manipulação, tratamento e análise dos dados.
- **Power BI:** Ferramenta de visualização de dados, utilizada para criar gráficos, dashboards e gerar relatórios interativos.
- **Planilhas do excel:** Usadas como base para organização, entrada e exportação dos dados necessários para o projeto.

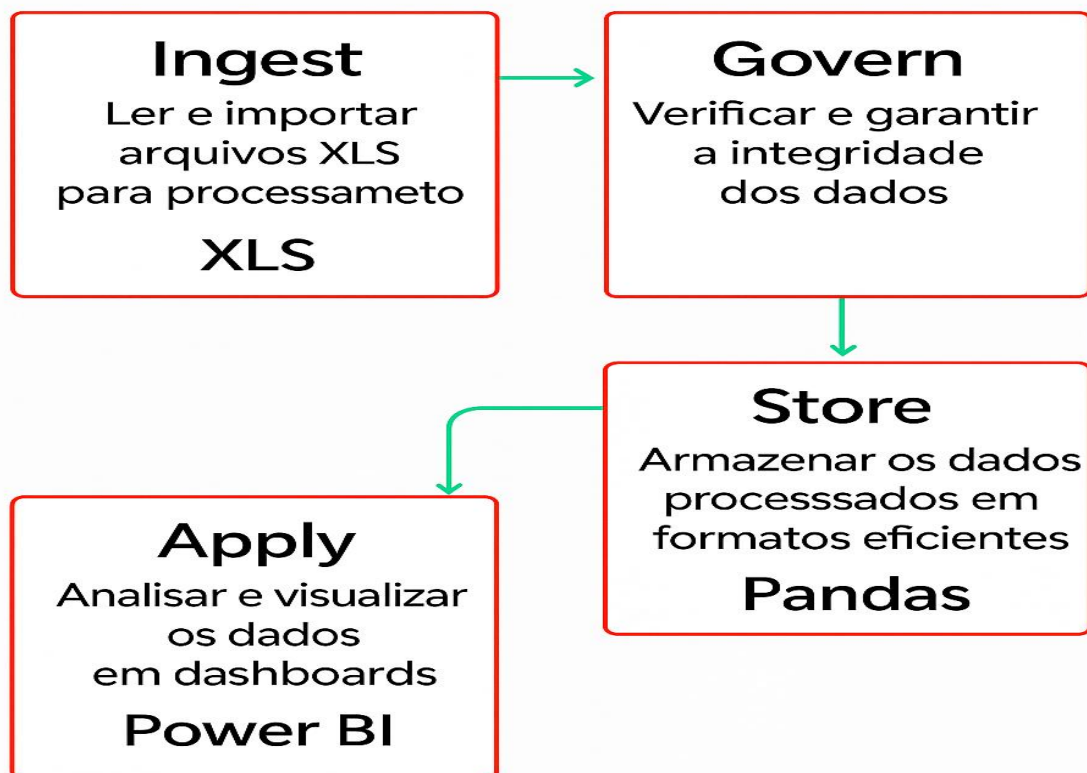
2.6. Detalhamento técnico do projeto

A solução consistiu no desenvolvimento de um sistema de análise e organização de dados para a empresa MateGuaraná, utilizando as linguagens e ferramentas Python, Pandas e Power BI. O projeto tem como foco principal automatizar o tratamento dos dados relacionados às vendas, produtos e desempenho da empresa, tornando o processo de análise mais rápido, confiável e eficiente.

Por meio do Python, aliado à biblioteca Pandas, foi possível desenvolver scripts para a limpeza, transformação e organização dos dados, corrigindo inconsistências, removendo duplicados e estruturando as informações de forma adequada para análise. Esses dados tratados foram então integrados ao Power BI, onde foram criados dashboards interativos e relatórios visuais, permitindo uma visualização clara e precisa dos indicadores mais importantes da empresa.

Essa solução permite que a MateGuaraná acompanhe de forma ágil informações como volume de vendas, controle de estoque, desempenho dos produtos e evolução dos resultados. Com isso, o processo de tomada de decisão se torna mais eficiente e estratégico, além de reduzir erros manuais no controle dos dados.

Arquitetura do Projeto – Big Data em Python



3. ENCERRAMENTO DO PROJETO

3.1. Relato Coletivo:

De acordo com os objetivos estabelecidos no subtópico 2.4, consegui atingir com sucesso as metas propostas. O projeto permitiu organizar e visualizar os dados de forma muito mais eficiente, facilitando a análise e interpretação das informações pela empresa. Durante o desenvolvimento, ampliei meu conhecimento no uso de ferramentas como Python, que foi essencial para o tratamento e organização dos dados; Pandas, que possibilitou a manipulação eficiente dos conjuntos de dados; Power BI, utilizado para criar dashboards e relatórios interativos e visuais; e também planilhas, que serviram como base para estruturar e importar os dados necessários.

De modo geral, os objetivos foram plenamente alcançados, e a experiência foi extremamente enriquecedora, tanto no aspecto técnico quanto na prática de resolver demandas reais, alinhadas ao que aprendi em sala de aula.

3.1.1. Avaliação de reação da parte interessada

As respostas não podem ser editadas

Entrevista com a parte interessada.

Entrevista com o público interessado no projeto com o intuito de obter feedback e uma avaliação sobre a ação realizada.

O e-mail do participante (elahyne.alves@gmail.com) foi registrado durante o envio deste formulário.

Email:

mateguarana1@gmail.com

Nome:

Mateguarana

Aspectos visuais do programa

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☒ 5

Facilidade no uso do programa

- ☐ 1
- ☐ 2
- ☐ 3
- ☒ 4
- ☐ 5

Utilidade operacional

- ☐ 1
- ☐ 2
- ☐ 3
- ☒ 4
- ☐ 5

Atendeu suas expectativas

- ☐ 1
- ☐ 2

Atendeu suas expectativas

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☒ 5

Nota geral do programa

- ☐ 1
- ☐ 2
- ☐ 3
- ☒ 4
- ☐ 5

O que poderia ser melhorado no programa?

Melhoria no excel

Enviada: 02/06/2025, 15:42

3.2. Relato de Experiência Individual (Pontuação específica para o relato individual)

Relato individual de Victor Costa de Oliveira:

3.2.1. CONTEXTUALIZAÇÃO

Atuei como único atuante do projeto, participando e realizando todas as etapas de maneira individual. Conversando com a empresa, tratando os dados, realizando os códigos e elaborando o PowerBI.

Este projeto me deu uma visão mais ampla de conversas com a parte interessada e metas, visto que, eu era quem definia minhas próprias metas do projeto. Me proporcionando uma experiência enriquecedora.

3.2.2. METODOLOGIA

A metodologia adotada para a execução deste projeto configurou-se como um estudo de caso aplicado, focado na resolução de um problema real da empresa MateGuaraná. O trabalho foi estruturado em fases sequenciais, desde o diagnóstico inicial até a entrega da solução final.

Fase 1: Diagnóstico e Levantamento de Requisitos

O ponto de partida foi o contato direto com a parte interessada para compreender suas operações e desafios. Foi identificado que a empresa realizava o controle de dados financeiros, de vendas e de estoque de forma manual, majoritariamente em planilhas. Essa prática gerava dificuldades na análise e na tomada de decisões estratégicas. O objetivo definido foi, então, transformar esses dados brutos em informações visuais e de fácil interpretação.

Fase 2: Coleta e Processamento de Dados (ETL)

Com os desafios mapeados, a fase seguinte concentrou-se no ciclo de Extração, Transformação e Carga (ETL) dos dados.

Extração: Os dados foram recebidos em formato de planilhas Excel, contendo informações sobre vendas, formas de pagamento e produtos. A extração desses dados foi realizada com o uso de scripts em Python, utilizando a biblioteca Pandas para ler e importar os arquivos para um ambiente de análise.

Transformação: Esta foi a etapa mais crítica. Utilizando a biblioteca Pandas, realizei um processo de limpeza e tratamento dos dados. As principais atividades incluíram:

A remoção de linhas vazias e dados duplicados para garantir a consistência.

A padronização de colunas, como as de datas, que apresentavam inconsistências de formato. O enriquecimento dos dados com a criação de novas colunas, como "Mês" e "Ticket Médio", para permitir análises mais aprofundadas e segmentadas.

Carga: Após o processamento, os dados limpos e estruturados foram carregados na ferramenta de visualização.

Fase 3: Visualização de Dados e Desenvolvimento dos Dashboards

Nesta fase, utilizei o Microsoft Power BI para desenvolver painéis interativos. Os dados tratados serviram como fonte para a criação de gráficos e relatórios visuais que permitem à empresa monitorar indicadores de desempenho, como:

Faturamento total e por forma de pagamento.

Desempenho de vendas ao longo dos meses.

Valores diários de lucro e ticket médio.

A solução visual foi projetada para ser intuitiva, permitindo que o gestor da empresa pudesse filtrar informações e obter insights de forma autônoma e eficiente.

Fase 4: Validação e Feedback

Ao final, a solução foi apresentada à parte interessada para validação. Por meio de um formulário de avaliação, a empresa forneceu um feedback sobre os aspectos visuais, a usabilidade e a utilidade da ferramenta, que atendeu às expectativas e demonstrou ter grande aplicabilidade operacional.

3.2.3. RESULTADOS E DISCUSSÃO:

Apreendi muito ao longo deste projeto e tive uma experiência extremamente positiva. Trabalhar com análise de dados utilizando Python, Pandas e Power BI exigiu uma compreensão sólida de conceitos como tratamento de dados, limpeza, organização e visualização de informações. Cada etapa do processo desde a interpretação das planilhas fornecidas pela empresa até o desenvolvimento dos dashboards apresentou desafios que exigiram foco, paciência e dedicação para serem superados.

Além disso, foi uma experiência nova entender como transformar grandes volumes de dados brutos em informações claras, organizadas e úteis para o negócio. Aprender a criar gráficos, filtros e relatórios interativos no Power BI, junto com os processamentos dos dados com Pandas, foi fundamental para entregar uma solução que realmente atendesse às necessidades do cliente.

Print com os códigos:

Mate guarana_code_1: Neste código, eu tentei unir tudo em uma única planilha, porém, o resultado não foi o esperado, ficando bastante confuso e com alguns problemas. Depois disto ser analisado, preferi adotar outro método, dividindo os códigos em partes, para facilitar a visualização e alteração (caso necessária)

```
1 import pandas as pd
2 import os
3 pasta_base = "C:/Users/202402369989/Desktop/mate guarana/"
4
5 arquivos = {
6     "venda_por_dia_jan": "VENDA_POR_DIA_JANEIRO_2025.csv",
7     "venda_por_produto_jan": "VENDA_POR_PRODUTO_JANEIRO_2025.csv",
8     "venda_por_dia_fev": "vendas_de_fevereiro.csv",
9     "venda_por_produto_fev": "vendas_de_fevereiro_por_produto.csv",
10    "venda_por_pagamento_fev": "vendas_de_fevereiro_por_pagamento.csv",
11    "faturamento_jan": "faturamento_JANEIRO_2025.csv",
12    "movimentacao_produtos_fev": "movimentacao_de_produtos_fevereiro.csv",
13    "planilha_custos": "PLANILHA_DE_CUSTO_2025_MATE_GUARANA.xlsx"
14 }
15
16 def carregar_csv(caminho):
17     try:
18         return pd.read_csv(caminho, sep=";", encoding="latin1")
19     except Exception as e:
20         print(f"Erro ao carregar {caminho}: {e}")
21         return pd.DataFrame()
22
23 dfs = {}
24 for nome, arquivo in arquivos.items():
25     caminho_completo = os.path.join(pasta_base, arquivo)
26     if not os.path.exists(caminho_completo):
27         print(f"X Arquivo não encontrado: {caminho_completo}")
28         continue
29
30     if caminho_completo.endswith(".csv"):
31         df = carregar_csv(caminho_completo)
32     else:
33         try:
34             df = pd.read_excel(caminho_completo)
35         except Exception as e:
36             print(f"Erro ao carregar {caminho_completo}: {e}")
37             df = pd.DataFrame()
38
39     if not df.empty:
40         dfs[nome] = df
41         print(f"\n✅ {nome.upper()} carregado com sucesso:")
42         print(df.head())
43     else:
44         print(f"\n⚠️ {nome} está vazio ou falhou ao carregar.")
45
46 tabelas_combinadas = []
47 for chave in ["venda_por_dia_jan", "venda_por_dia_fev", "venda_por_produto_jan", "venda_por_produto_fev"]:
48     if chave in dfs:
49         tabelas_combinadas.append(dfs[chave])
50
51 if tabelas_combinadas:
52     df_final = pd.concat(tabelas_combinadas, ignore_index=True)
53     output_path = os.path.join(pasta_base, "dados_tratados.csv")
54     df_final.to_csv(output_path, index=False)
55     print(f"\n✅ Arquivo final salvo em: {output_path}")
56 else:
57     print("\n⚠️ Nenhuma tabela foi combinada.")
58
```


Mateguarana_code_2: Neste código, eu separei os dados para visualizar APENAS os dados de venda do mês de Janeiro e Fevereiro, sem muitos problemas, consegui chegar ao resultado.

```
1 import pandas as pd
2 import os
3
4 pasta_base = "C:/Users/202402369989/Desktop/mateguarana/"
5
6 # dicionário com os nomes dos arquivos
7 arquivos = {
8     "venda_por_dia_jan": "VENDA_POR_DIA_JANEIRO_2025.csv",
9     "venda_por_produto_jan": "VENDA_POR_PRODUTO_JANEIRO_2025.csv",
10    "venda_por_dia_fev": "vendas_de_fevereiro.csv",
11    "venda_por_produto_fev": "vendas_de_fevereiro_por_produto.csv",
12    "venda_por_pagamento_fev": "vendas_de_fevereiro_por_pagamento.csv",
13    "faturamento_jan": "faturamento_JANEIRO_2025.csv",
14    "movimentacao_produtos_fev": "movimentacao_de_produtos_fevereiro.csv",
15    "planilha_custos": "PLANILHA_DE_CUSTO_2025_MATE_GUARANA.csv"
16 }
17
18 # carregar csv
19 def carregar_csv(caminho):
20     try:
21         return pd.read_csv(caminho, sep=";", encoding="latin1")
22     except Exception as e:
23         print(f"Erro ao carregar {caminho}: {e}")
24         return pd.DataFrame()
25
26 # carregar dicionario
27 dfs = {}
28 for nome, arquivo in arquivos.items():
29     caminho_completo = os.path.join(pasta_base, arquivo)
30     if caminho_completo.lower().endswith(".csv"):
31         df = carregar_csv(caminho_completo)
32     else:
33         try:
34             df = pd.read_excel(caminho_completo)
35         except Exception as e:
36             print(f"Erro ao carregar {caminho_completo}: {e}")
37             df = pd.DataFrame()
38
39     if not df.empty:
40         dfs[nome] = df
41         print(f"✅ {nome.upper()} carregado com sucesso.")
42     else:
43         print(f"⚠️ {nome} está vazio ou falhou ao carregar.")
44
45 # verificar se foi carregado
46 df_jan_dia = dfs.get("venda_por_dia_jan")
47 df_fev_dia = dfs.get("venda_por_dia_fev")
48
49 # exportar limpos
50 if df_jan_dia is not None:
51     caminho_jan_corrigido = os.path.join(pasta_base, "dados_janeiro.csv")
52     df_jan_dia.to_csv(caminho_jan_corrigido, index=False, encoding="utf-8")
53     print(f"📁 Arquivo de janeiro exportado: {caminho_jan_corrigido}")
54
55 if df_fev_dia is not None:
56     caminho_fev_corrigido = os.path.join(pasta_base, "dados_fevereiro.csv")
57     df_fev_dia.to_csv(caminho_fev_corrigido, index=False, encoding="utf-8")
58     print(f"📁 Arquivo de fevereiro exportado: {caminho_fev_corrigido}")
59
```

Mateguarana_code_3: Neste código, eu separei os pagamento, para ver tanto os métodos de pagamento utilizados nos meses, quanto os produtos vendidos e seus valores, também sem muitos problemas encontrados nesta fase.

```
1 import pandas as pd
2 import os
3
4 # caminho base
5 pasta_base = "C:/Users/202402369989/Desktop/mateguarana"
6
7 # arquivos
8 arquivos = {
9     "vendas_jan": os.path.join(pasta_base, "VENDA_POR_DIA_JANEIRO_2025.csv"),
10    "vendas_fev": os.path.join(pasta_base, "vendas_de_fevereiro.csv"),
11    "vendas_prod_jan": os.path.join(pasta_base, "VENDA_POR_PRODUTO_JANEIRO_2025.csv"),
12    "vendas_prod_fev": os.path.join(pasta_base, "vendas_de_fevereiro_por_produto.csv"),
13    "pagamento_fev": os.path.join(pasta_base, "vendas_de_fevereiro_por_pagamento.csv"),
14    "pagamento_jan": os.path.join(pasta_base, "pagamentos_janeiro.csv"),
15 }
16
17 def carregar_csv(caminho, sep=";", encoding="latin1"):
18     return pd.read_csv(caminho, sep=sep, encoding=encoding)
19
20 def limpar_espacos(series):
21     return series.astype(str).str.strip()
22
23 def formatar_valores(series):
24     def formatar(v):
25         try:
26             valor = float(
27                 str(v).replace("R$", "").replace(".", "").replace(",", ".").strip()
28             )
29             return f"R$ {valor:,.2f}".replace(",", "v").replace(".", ",").replace("v", ".")
30         except:
31             return v
32     return series.apply(formatar)
33
34 # carregar arquivos
35 vendas_jan = carregar_csv(arquivos["vendas_jan"])
36 vendas_fev = carregar_csv(arquivos["vendas_fev"])
37 vendas_prod_jan = carregar_csv(arquivos["vendas_prod_jan"])
38 vendas_prod_fev = carregar_csv(arquivos["vendas_prod_fev"])
39 pagamento_fev = carregar_csv(arquivos["pagamento_fev"])
40 pagamento_jan = carregar_csv(arquivos["pagamento_jan"])
41
42 # processamento e renomeação de colunas
43 def processar_vendas(dia_df, prod_df, mes):
44     dia_df["Mês"] = mes
45     dia_df["Dia"] = dia_df["Dia"].str.strip()
46     prod_df = prod_df.rename(columns={"Descrição": "Produto", "Quant.Vendas": "Quantidade", "Valor Pago": "ValorPago"})
47     prod_df["Mês"] = mes
48     return prod_df
49
50 vendas_prod_jan_df = processar_vendas(vendas_jan, vendas_prod_jan, "Janeiro")
51 vendas_prod_fev_df = processar_vendas(vendas_fev, vendas_prod_fev, "Fevereiro")
52
53 vendas_produtos_total = pd.concat([vendas_prod_jan_df, vendas_prod_fev_df], ignore_index=True)
54
55 for col in ["Quantidade", "ValorPago", "Custo Total"]:
56     if col in vendas_produtos_total.columns:
57         vendas_produtos_total[col] = formatar_valores(vendas_produtos_total[col])
58
59 pagamento_fev["Meio de pagamento"] = limpar_espacos(pagamento_fev["Meio de pagamento"])
60 pagamento_fev["Total"] = formatar_valores(pagamento_fev["Total"])
61 pagamento_fev["Mês"] = "Fevereiro"
62
63 pagamento_jan["Meio de pagamento"] = limpar_espacos(pagamento_jan["Meio de pagamento"])
64 pagamento_jan["Total"] = formatar_valores(pagamento_jan["Total"])
65 pagamento_jan["Mês"] = "Janeiro"
66
67 pagamentos = pd.concat(
68     [pagamento_jan[["Meio de pagamento", "Total", "Mês"]],
69     pagamento_fev[["Meio de pagamento", "Total", "Mês"]]],
70     ignore_index=True
71 )
72
73 # exportação
74 vendas_produtos_total.to_csv("vendas_produtos_integradas.csv", index=False, encoding="utf-8-sig")
75 pagamentos.to_csv("pagamentos_integrados.csv", index=False, encoding="utf-8-sig")
76
77 print("✅ Integração concluída. Arquivos exportados:")
78 print("- vendas_produtos_integradas.csv")
79 print("- pagamentos_integrados.csv")
80
```

Mate guarana_projecao_marco: Nesta parte, eu fiz uma projeção para o mês de Março, tendo em mente que Março é um mês que tem um aumento no fluxo de vendas por conta do carnaval, então coloquei um aumento de 50% nas vendas neste mês.

```
1 import pandas as pd
2
3 jan_path = "C:/Users/202402369989/Desktop/mate guarana_projeto/mate guarana_projeto/mate guarana_planilhas/VENDA POR PRODUTO JANEIRO 2025.csv"
4 fev_path = "C:/Users/202402369989/Desktop/mate guarana_projeto/mate guarana_projeto/mate guarana_planilhas/vendas de fevereiro por produto.csv"
5 ref_path = "C:/Users/202402369989/Desktop/mate guarana_projeto/mate guarana_projeto/codigos_mate guarana_com_planilhas/mate guarana_code_3/vendas_produtos_integradas.csv"
6
7 # ler csv
8 jan_df = pd.read_csv(jan_path, sep=";", encoding="utf-8-sig")
9 fev_df = pd.read_csv(fev_path, sep=";", encoding="utf-8-sig")
10 ref_df = pd.read_csv(ref_path, sep=";", encoding="utf-8-sig")
11
12 # rename primeira coluna como produto
13 jan_df = jan_df.rename(columns={jan_df.columns[0]: "Produto"})
14 fev_df = fev_df.rename(columns={fev_df.columns[0]: "Produto"})
15 ref_df = ref_df.rename(columns={ref_df.columns[0]: "Produto"})
16
17 # padronizar nome de quantidade
18 for df in [jan_df, fev_df]:
19     if "Quant.Vendas" in df.columns:
20         df.rename(columns={"Quant.Vendas": "Quantidade"}, inplace=True)
21
22 # converter colunas monetárias para float
23 def parse_monetario(valor):
24     try:
25         return float(str(valor).replace("R$", "").replace(".", "").replace(",", ".").strip())
26     except:
27         return 0.0
28
29 monetarias = ["Valor Pago", "Total Vendas", "Custo Total", "ValorPago"]
30
31 for col in monetarias:
32     if col in jan_df.columns:
33         jan_df[col] = jan_df[col].apply(parse_monetario)
34     if col in fev_df.columns:
35         fev_df[col] = fev_df[col].apply(parse_monetario)
36     if col in ref_df.columns:
37         ref_df[col] = ref_df[col].apply(parse_monetario)
38
39 # problema com nomes
40 if "ValorPago" in ref_df.columns:
41     ref_df = ref_df.rename(columns={"ValorPago": "Valor Pago"})
42
43 # quantidades como numero
44 for df in [jan_df, fev_df, ref_df]:
45     df["Quantidade"] = pd.to_numeric(df["Quantidade"], errors="coerce").fillna(0)
46
47 # groupby por produto e somar colunas relevantes
48 jan_group = jan_df.groupby("Produto")[["Quantidade", "Valor Pago", "Total Vendas", "Custo Total"]].sum()
49 fev_group = fev_df.groupby("Produto")[["Quantidade", "Valor Pago", "Total Vendas", "Custo Total"]].sum()
50 ref_group = ref_df.groupby("Produto")[["Quantidade", "Valor Pago", "Total Vendas", "Custo Total"]].sum()
51
52 # media
53 media_df = (jan_group + fev_group) / 2
54
55 # projecao
56 marco_proj = media_df * 1.5
57
58 # limitar em 50K
59 for col in marco_proj.columns:
60     if col in ref_group.columns:
61         limite = ref_group[col] * 0.5
62         marco_proj[col] = marco_proj[col].where(marco_proj[col] <= limite, limite)
63
64 # resetar index e formatar resultado
65 marco_proj = marco_proj.reset_index()
66 marco_proj["Mes"] = "Março"
67 marco_proj["Quantidade"] = marco_proj["Quantidade"].fillna(0).round(0).astype(int)
68
69 # passar valores para reais novamente
70 def formatar_rs(valor):
71     return f"R$ {valor:,.2f}".replace(".", "").replace(",", ".").replace("X", ".")
72
73 for col in ["Valor Pago", "Total Vendas", "Custo Total"]:
74     marco_proj[col] = marco_proj[col].apply(formatar_rs)
75
76 # exportar tudo
77 marco_proj.to_csv("projecao_marco.csv", index=False, encoding="utf-8-sig")
78 print("✅ Arquivo 'projecao_marco_final.csv' gerado com sucesso!")
79
```

3.2.4. REFLEXÃO APROFUNDADA

Minha participação no projeto foi uma experiência valiosa, que me permitiu transformar o conhecimento aprendido em sala de aula em soluções reais. Percebi, na prática, como o trabalho em equipe e o entendimento das demandas do cliente são fundamentais para alcançar bons resultados. Esse projeto me mostrou, de forma clara, o quanto a tecnologia e a análise de dados podem ser ferramentas poderosas para otimizar processos e gerar valor para os negócios.

3.2.5. CONSIDERAÇÕES FINAIS

O projeto com certeza foi enriquecedor, durante o projeto, identifiquei que ainda existem outros aspectos que podem ser explorados futuramente junto à parte interessada, como a automação na atualização dos dados, a criação de alertas inteligentes para controle de estoque, ou até mesmo a implementação de modelos preditivos para estimar vendas e demandas futuras, tornando a gestão ainda mais eficiente.

Do ponto de vista de extensão e pesquisa, seria possível expandir o projeto com o desenvolvimento de um sistema próprio para gestão de vendas e estoque, integrando diretamente os dados em tempo real, eliminando a dependência de planilhas manuais. Também poderiam ser realizados estudos mais aprofundados sobre o comportamento de compra dos clientes.