**TESTE TÉCNICO - ANALISTA DE DADOS - MIND LAB**

**Dinâmica das pastas**

**Da pasta raiz “Desafio Técnico” temos:**

* **Desafio Técnico\Manipulação de Dados em Excel e Google Sheets**
  + **PARTE 1 - Manipulação de Dados em Excel e Google Sheets:**
* **Desafio Técnico\Visualização de dados em Power BI**
  + **PARTE 2 - Visualização de dados em Power BI:**
* **Desafio Técnico\Automação com Python**
  + **PARTE 3 - Automação com Python:**

**PARTE 1 - Manipulação de Dados em Excel e Google Sheets:**

Link do google sheets: <https://docs.google.com/spreadsheets/d/1KAH8Bw3KWlBADUc19VmVv1EjeNpONDl1uXU-G3XJP7Y/edit?gid=1005822868#gid=1005822868>

Foram feitos ajustes para as colunas na aba “dados\_vendas\_2023.csv”:

* **Produto ID:** As células como valor “DEZ” foram substituídas pelo valor “10” para padronizar o ID. Os valores foram postos em uam nova coluna denominada “Produto ID Ajustado”.
  + **Formula:** =SE(B2="DEZ";10;B2)
* **Região Ajustada:** Os nomes foram padronizados colocando em caixa alta e substituindo o valor “CENTROESTE” pelo valor "CENTRO-OESTE". Os valores foram postos em uam nova coluna denominada “Região Ajustada”.
  + **Formula:** =SE(MAIÚSCULA(C2)="CENTROESTE";"CENTRO-OESTE";MAIÚSCULA(C2))
* **Valor:** As strings numéricastiveram os caracteres de virgula substituídos pelo ponto e depois convertidas para número. As linhas que não puderam ser convertidas ficaram como null. Os valores foram postos em uam nova coluna denominada “Valor Ajustado”.
  + **Formula:** =SE(ÉNÚM(D2);D2;)
* **Quantidade:** As células que não eram numéricas ficaram como vazio. Os valores foram postos em uam nova coluna denominada “Quantidade Ajustado”.
  + **Formula:** =SE(ÉNÚM(D2);D2;)

Os cálculos de vendas totais por produto e por região estão disponíveis na aba "Informações sobre Vendas Totais". Para realizar esses cálculos, utilizamos a função FILTER, que filtra as linhas de acordo com o produto e a região selecionados.

As fórmulas não necessitam de ajustes, pois o tratamento dos dados já foi previamente realizado. As listas suspensas de produtos e regiões são extraídas diretamente da base de dados e são atualizadas automaticamente conforme a base é modificada. Além disso, as referências no FILTER abrangem toda a coluna, garantindo que novas linhas sejam consideradas à medida que a base de dados cresce.

A tabela dinâmica foi adicionada usando a própria ferramenta nativa do google sheets e está na aba “Tabela dinâmica sobre vendas totais”.

Para adaptar o fluxo de trabalho para o Excel basta adaptar a nomenclatura de algumas formulas e a substituição de outras como o FILTER pelo SOMASES. No mais, o processo funcionaria de forma muito similar.

**PARTE 2 - Visualização de dados em Power BI:**

O arquivo do Power BI foi nomeado como “Desafio Técnico.pbix”. Foram feitos ajustes na base para remoção de duplicatas, espaços vazios a mais, correção de termos e outros ajustes menores para garantir a homogeneidade da base.

Em seguida foi feito o relacionamento dos dados usando as colunas Produto ID na escala de (1;\*) da base produtos para a base dados\_vendas, uma vez que temos repetição de Produto ID na base dados\_vendas e não na base produtos.

Em seguida foram gerados os gráficos Vendas totais por mês; Vendas por região; Vendas por tipo de produto; além do card de Total de vendas acumuladas. O dashboard foi dividido em duas páginas, uma pegando os valores de venda (R$) “Dashboard - Valores Vendas” e outro com a quantidade de vendas “Dashboard - Quantidade de Vendas”.

Já o cálculo mês a mês não foi desenvolvido, pois não tive tempo hábil para encontrar o erro que estava dando na fórmula então priorizei entregar outros tópicos do desafio.

**PARTE 3 - Automação com Python:**

O script foi dividido em um script para tratamento dos dados “Script\_Tratamento\_Base\_De\_Dados.py” e outros para análise “Script\_Análise\_Base\_De\_Dados.py”. Ao executar o script de tratamento gera-se o arquivo “Base\_dados\_financeiros\_tratada.csv” que é utilizado no script de análise.

A forma como o script funciona está documentado no próprio arquivo. Como resultado analítico observou-se uma elevada variabilidade do dos dados (desvio padrão) e não foi observada nenhuma tendência seja para valor, quantidade e valor por unidade na análise temporal (análsie gráfica). Por falta de tempo o Webscraping também não foi realizado.