

# Challenge Data Engineer DWE 2024

## 3ª EDIÇÃO DATAEX BOOTCAMP DATA WOMEN ENGINEERS



### Objetivo:

Avaliar conhecimentos adquiridos no DataEX Bootcamp Data Women Engineers 2024

### Público-alvo:

Alunas do DataEX Bootcamp Data Women Engineers 2024 (3ª Edição)

### O Desafio irá testar suas habilidades nas seguintes competências:

- ✓ Construção de 1 Pipeline de dados com Data factory (Pipeline) e Notebook no Fabric (Max 700 Pts).
- ✓ Habilidades com a linguagem SQL Comandos DML (Max 300 Pts).

## Termos e condições:

- Ser aluna inscrita e ativa no DataEX Bootcamp Data Women Engineers 2024;
- Não ser colaborador ou prestador de serviço da DataEX.

## Para tornar os prêmios deste desafio elegíveis, as participantes devem:

- Entregar o Desafio até a data final: **18/02/2024 às 23:59h;**
- O arquivo deverá ser enviado única e exclusivamente para o e-mail [treinamentos@dataex.com.br](mailto:treinamentos@dataex.com.br).

Ao participar do Challenge Data Engineer do DataEX Bootcamp Data Women Engineers 2024 - 3ª Edição, eu AUTORIZO o uso da minha imagem e vídeo contidos no Bootcamp DWE, realizado do dia 29/01/2024 ao dia 08/02/2024 em todo e qualquer material impresso e/ou virtual, como sites, redes sociais, reportagens, vídeos institucionais, propaganda e outros desenvolvidos pela DataEX, sejam essas destinadas à divulgação ao público em geral.

A presente autorização é concedida a título gratuito, abrangendo o uso da imagem acima mencionada em todo território nacional e no exterior, sem data de expiração. Por esta ser a expressão da minha vontade, declaro que autorizo o uso acima descrito sem que nada haja a ser reclamado a título de direitos conexos à minha imagem ou a qualquer outro, e assino a presente autorização.

## Prazo Final da entrega do Desafio:

18/02/2024 às 23:59h (Domingo)

**Desafios enviados após a data descrita serão desclassificados automaticamente.**

## Critérios de desempate

Será pela ordem de entrega, priorizando ordem de recebimento.

## Forma de envio:

*Entrega Pacote único zipado e nomeado com seu usuário de banco de dados no Bootcamp:*

*Exemplo: DWEXX.zip*

***Organize todos os arquivos criados da melhor forma possível e cronológica para avaliação.***

*Enviar para e-mail [treinamentos@dataex.com.br](mailto:treinamentos@dataex.com.br)*

Assunto E-mail "Entrega Desafio DWE 2024 DWEXX"

Assunto E-mail "Entrega Desafio DWE 2024 DWE99"

## Divulgação dos Resultados

28/02/2024 às 15h (Live no Sympla)

## Premiação Challenge DWE 2024

3º Lugar



2º Lugar



1º Lugar



OBS: Parte 1 e 2, possuem dependência entre elas em ordem cronológica.

# 1ª Parte – Desenvolvimento Pipeline/Notebook Visual Code

## Requisitos de Negócio

Vale até 700 pontos

### O que vamos analisar?

O Arquivo do tipo CSV com Assuntos Vendas com o nome de Vendas.csv

### Como desejamos analisar os dados?

- Produto
- Categoria
- Segmento
- Fabricante
- Cliente (Nome)
- Cidade
- Estado
- Região
- País

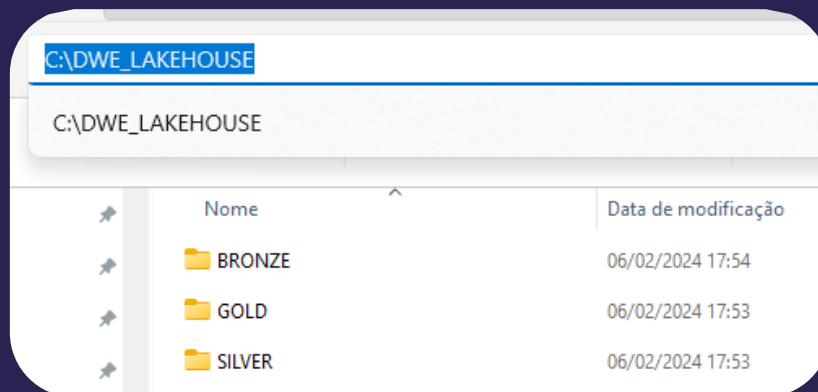
### Quais métricas desejamos acompanhar?

- Total Vendas (R\$)
- Total Custo (R\$)
- Total Lucro (R\$)

- Total Quantidade (Unidades)

## Requisitos de Técnicos

Desenvolva um o Notebook no Visual Code e sua máquina local e realize a modelagem Star Schema, desnormalizando os dados gerando um fluxo de dados de acordo com seu entendimento do escopo do desafio e realize as transformações de dados usando a estrutura de pastas da arquitetura Medallion, conforme abaixo.



- Crie seu Notebook com o nome DESAFIO\_DWEXX.ipynb (XX Colocar seu número de Bootcamper)
- Gere scripts com possibilidade de reprocessamentos.
- Gere um diagrama dimensional conceitual com a modelagem aplicada. (Adicione a imagem de sua modelagem em seu notebook)
- Use nomes intuitivos em seus objetos.
- Use Linguagem Python com a Biblioteca Pandas
- Camadas Medallion
  - Bronze: Contém o arquivo disponibilizado;
  - Silver: Contém arquivos transformados em formato parquet;
    - Tratar a coluna nome, separando e-mail de nome
    - Tratar a coluna cidade, mantendo apenas o nome da cidade.
    - Eliminar as colunas IDCampanha, distrito e código postal
  - Gold: Contém os arquivos prontos para uso em formato parquet;
    - Adicionar as SK's as tabelas dimensões.
    - Aplicar SCD1 nas dimensões.
    - Adicionar as SK's tabela fato.
    - Particionar os arquivos Parquet da tabela fato em ano e mês
- Importe os arquivos Parquet da camada GOLD no Power BI
  - Configure o diagrama dimensional Lógico no Power BI
  - Gere uma tabela calendário no DAX do Power BI.

## Origem de dados:

Nome do Arquivo	Assunto
Vendas.csv	Vendas

## 2ª Parte – Conhecimentos PYTHON PANDAS

DESENVOLVA SCRIPTS/BLOCOS EM SEU NOTEBOOK PARA CADA PERGUNTA DE NEGÓCIO ABAIXO COM PYTHON PANDAS:

### EXERCÍCIO 1

*Vale até 30 pontos*

Qual o total de Vendas em Fev/2014 para cada Fabricante?

**ENTREGÁVEL: SCRIPT NO NOTEBOOK DO DESAFIO**

### EXERCÍCIO 2

*Vale até 30 pontos*

Quais os top 10 Produtos referente a Jan/2014 em Total de custo (R\$)

**ENTREGÁVEL: SCRIPT NO NOTEBOOK DO DESAFIO**

### EXERCÍCIO 3

*Vale até 30 pontos*

Qual o total de vendas no 1º Trimestre/2014 por Segmento?

**ENTREGÁVEL: SCRIPT NO NOTEBOOK DO DESAFIO**

### EXERCÍCIO 4

*Vale até 30 pontos*

Quais as Top 5 Segmentos em 2014 sobre o total de Vendas(R\$)?

**ENTREGÁVEL: SCRIPT NO NOTEBOOK DO DESAFIO**

### EXERCÍCIO 5

*Vale até 30 pontos*

Qual o Total de Vendas(R\$) mês a mês em 2014?

**ENTREGÁVEL: SCRIPT NO NOTEBOOK DO DESAFIO**

### EXERCÍCIO 6

*Vale até 30 pontos*

Qual o total de Lucro mês a mês em 2014?

**ENTREGÁVEL: SCRIPT NO NOTEBOOK DO DESAFIO**

### EXERCÍCIO 7

*Vale até 30 pontos*

Qual a variação %(MoM) de Total de vendas (R\$) de Mar/2014 para Fev/2014 por Segmento.

**ENTREGÁVEL: SCRIPT NO NOTEBOOK DO DESAFIO**

### EXERCÍCIO 8

*Vale até 30 pontos*

Retorne o maior Valor de custos (R\$) e seu respectivo segmento em Fev/2014.

**ENTREGÁVEL: SCRIPT NO NOTEBOOK DO DESAFIO**

### EXERCÍCIO 9

*Vale até 30 pontos*

Retorne o % (Share) de Total de vendas (R\$), por Segmento em Fev/2014.

**ENTREGÁVEL: SCRIPT NO NOTEBOOK DO DESAFIO**

### EXERCÍCIO 10

*Vale até 30 pontos*

Retorne o valor médio de Total de vendas Mês a Mês por segmento no ano de 2014.

**ENTREGÁVEL: SCRIPT NO NOTEBOOK DO DESAFIO.**

## ORIENTAÇÕES FINAIS

- Entregar notebook único com parte 1 e 2 (Ordem cronológica).
- Entregar o arquivo PBIX com arquivo parquet com nome DWEXX.pbix

Compactar o Notebook desenvolvido e o arquivo PBIX entregar apenas um arquivo ZIP com o nome de *Ex: DWEXX.zip*

- Organize e evidencie a construção do seu pipeline da melhor forma possível para entrega e avaliação.
- Documente em seu Notebook todos os passos realizados.

**ENTREGUE O MÁXIMO POSSÍVEL DE EXERCÍCIOS.  
SERÃO ACEITAS ENTREGAS PARCIAIS.**

# Boa sorte!

Deseja ir além do solicitado? Vá além!  
Agora é com você! Bora codar, Bootcampers!  
Baixe os arquivos no Sympla e Hands-on!