**DESAFIO FINAL - BD/BI - 2025.1**

**FÁBRICA DE SOFTWARE**

**POWER BI**

1 – Baixe e importe um dataset de alguns desses sites: **kaggle, dados.gov.br, UCI Machine Learning Repository**. Se preferir, use qualquer outro dataset (menos o que foi visto em sala e os que foram passados para o desafio diário) que tenha pelo menos 500 linhas;

2 – No Power BI, **atenção** à tipagem das colunas do seu dataset, e aos **erros** (valores nulos, duplicatas, etc.) que o dataset pode apresentar;

3 – Faça no **mínimo** 3 gráficos de tipos diferentes (Exemplo: Barras, Colunas, Pizza, Treemap, etc.) a partir do seu dataset. Atente-se no visual, cuidado na escolha de cores e efeitos;

4 – Com o Dashboard pronto, documente sobre todo o tratamento de dados que foi feito, e discorra sobre as conclusões da sua análise em um **arquivo .txt**. Por exemplo: “A cidade x teve uma média de vendas maior que a cidade y, isso pode ocorrer devido a...”;

5 – Sinta-se à vontade para analisar qualquer elemento do dataset e criar quantos gráficos e páginas quiser. Importante mostrar sua **capacidade de análise,** e **extrair insights** do dashboard;

OBS: Envie os 3 arquivos: power BI (.pbix), dataset (.xlsx ou .csv), e o texto (.txt) ou Word (.docx).

**Modelagem – SQL**

**Proposta:** Aumentar a complexidade da loja de departamento (editar o desafio feito no workshop de modelagem)

1. Modelagem de dados: conceitualização e ferramental.

1.1. Editar o banco de dados composto por no mínimo 05 entidades(tabelas) com 03 ou 04 atributos(colunas);

1.2. Apresentar a modelagem conceitual (ER), lógica e física do banco de dados;

1.3. Construir (ou manter as já existentes) pelo menos 03 relações de herança de atributos entre as tabelas [Chave estrangeira]

1.4. Descrever os relacionamentos existentes entre as entidades do modelo, identificando qual a chave-primária, sua cardinalidade e qual a tabela fato e dimensão do relacionamento.

2. DDL e DML: definição e manipulação de dados.

2.1. Conectar ao servidor do MySQL Workbench;

2.2. Construir um esquema que represente o banco de dados criado do modelo;

2.3. Construir as tabelas que representam as entidades do modelo;

2.4. Atualizar as tabelas com as especificações dos atributos solicitados no item 1.1;

2.5. Insira de 8 a 10 linhas por tabela.

3. DQL: consultas e operações de agregação e agrupamentos.

3.1. Realizar 01 consulta envolvendo ações de agregação ou agrupamentos **para cada tabela** [Variar as funções e atributos consultados];

3.2. Realizar 02 operações de JOIN **entre as tabelas** [O tipo de JOIN é critério do participante];

3.3. Descrever as informações adquiridas nas consultas e o comportamento dos dados consultados [Focar na interpretação dos dados e geração de insight].

Obs. 01: Salve as operações realizadas em formato .sql e suba na sua pasta do GitHub.

Obs. 02: Será necessário que exista um sentido lógico entre as tabelas, atributos e os dados inseridos.

Obs. 03: Para os modelos conceituais e lógicos, suba os arquivos do brmodelo e os prints de tela, para o modelo físico copiar para um arquivo txt ou pdf.

**PYTHON**

1 - Banco de Dados

Para execução do desafio, você deve criar um notebook (.ipynb), importar as bibliotecas Pandas e Matplotlib, e o banco de dados do link abaixo:

Link: <https://drive.google.com/drive/folders/1LTYnIkr9f12f-rCvo4JgAumbPbinB8Ep?usp=share_link>

2 - Tasks Finais

A ideia é que você faça uma Análise Completa dos Dados, desde visualizar suas características, a qualidade de seus dados e extrair valor dele através da geração de gráficos. Com isso, as questões propostas são:

1. **Características do Banco de Dados**
   1. Quantas linhas e colunas temos no banco de dados?
   2. Existem dados nulos no banco? Se sim, quais colunas apresentam e quantos são os dados nulos nelas?
   3. Quais tipos de dados estão presentes no banco de dados? (Exemplo: object, int, float)
2. **Agrupamento de Dados e Plotagem de Gráficos**

Utilizando as técnicas de agrupamento de dados vistas em sala, faça as seguintes questões:

1. Plote um gráfico de barras (bar) que exiba a quantidade de funcionários em cada setor (coluna Department);
2. Plote um gráfico de linha (line) que exiba a quantidade de funcionários para cada nível de escolaridade (coluna Education);
3. Plote um gráfico de torta/pizza (pie) que exiba a quantidade de funcionários que tem ou não algum atrito na empresa (coluna Attrition).

|  |
| --- |
| **ENTREGA**  Compartilhe o link da pasta ou repositório com o nome DESAFIO-FS-BDBI no grupo do whatsapp.  A pontuação do imersionista será computada de acordo com a conformidade das respostas e a postagem da atividade na data prevista no Github e no grupo.  **Horário limite para entrega do material: até as 12:00h do dia 02/03.** |