

Case Prático - Estágio em Engenharia de Analytics (Banco Shield)

Universo Marvel | Dados fictícios | Entrega via Git

1. Enredo

No multiverso da Marvel, a S.H.I.E.L.D. decidiu criar um braço bancário para financiar tecnologia, mobilidade e proteção das pessoas: o Banco Shield. Do outro lado, a organização Hidra vem crescendo silenciosamente e ganhando participação em produtos financeiros, operando com uma agressividade comercial que preocupa a diretoria.

Na última reunião do Conselho (com participação especial de Nick Fury e da equipe de analytics), a pergunta foi direta: "Onde estamos ganhando, onde estamos perdendo e onde o risco pode explodir?"

Você foi convocado(a) para apoiar o time de Engenharia de Analytics do Banco Shield. Seu objetivo é transformar dados brutos em informação confiável e açãoável para orientar uma decisão de negócio.

2. O problema de negócio

Hoje, a área de negócio não consegue responder com segurança a perguntas básicas, porque:

- Os dados chegam pouco tratados e inconsistentes, dificultando comparações e cortes por produto, localidade e concorrência.
- Há baixa confiança: uma pequena fração dos registros está incompleta ou incorreta, e isso contamina indicadores (produção, risco e oportunidade).
- Mesmo quando os números existem, faltam análises com narrativa e recomendações claras (o que fazer a partir disso).
- Não existe um painel de acompanhamento padronizado para o negócio monitorar evolução e resultados.

3. Perguntas-chave que a diretoria quer responder

1. Quais produtos são os mais vendidos (em unidades e em valor) no Banco Shield e na Hidra?
2. Em quais localidades/regiões cada banco é mais forte?
3. Quais produtos e localidades concentram maior risco de inadimplência (ex.: 30+ DPD) e qual o trade-off com volume?

4. Qual é a maior oportunidade de crescimento para o Banco Shield: (i) ganhar share onde já é forte, (ii) atacar nichos onde a Hidra domina, ou (iii) reduzir risco para liberar capacidade de crescimento?
5. Quais sinais de qualidade de dados precisam virar controles automáticos (ex.: chaves inválidas, campos vazios, valores negativos, duplicidades) para garantir SLA e confiabilidade?

4. Dados disponíveis

Você receberá três arquivos (CSV/TXT), com uma tabela fato e duas dimensões:

- fato_contratos.csv (fato): contratos/ocorrências por mês, banco, produto e localidade, incluindo métricas de volume, saldo e inadimplência.
- dim_produto.csv (dimensão): catálogo de produtos (nome, categoria, prazo e taxa base).
- dim_localidade.csv (dimensão): catálogo de localidades (nome, macro-região e fator de risco regional).
- metadados.txt: dicionário de dados e regras esperadas.

Importante: trate os dados como um cenário real. Você não receberá uma lista explícita de erros. Parte do exercício é identificar e propor como controlar esses problemas.

5. O que esperamos na entrega

Ao final, esperamos que você entregue (via Git) os seguintes itens:

- Tratamento dos dados: limpeza, padronização e preparação para análises e consumo em dashboard.
- Avaliação de qualidade: regras, checagens e evidências do que foi encontrado (incluindo proposta de controles automáticos).
- Análise para o negócio: um PPT com storyline (perguntas, abordagem, achados, recomendações).
- Dashboard de acompanhamento: painel com os principais KPIs e cortes (produto, localidade, banco), permitindo monitorar volume, share e risco.

Você pode usar Python e/ou SQL (recomendado). Para o dashboard, escolha a ferramenta com a qual você se sente mais confortável (ex.: Power BI, QuickSight, Looker Studio, Streamlit, Excel etc.).

6. Como entregar via Git (passo a passo resumido)

6. Acesse github.com e clique em Sign up (criar conta gratuita).
7. Crie seu usuário, confirme e-mail e finalize o cadastro.
8. Clique em New repository e crie um repositório (público). Sugestão de nome: case-banco-shield.

9. Faça upload do seu projeto (código, dados tratados/outputs e documentação). Você pode subir via interface web ou via Git no seu computador.
10. Caso ache importante, inclua um README.md com: como rodar seu pipeline, quais decisões tomou, e como acessar o dashboard (link, arquivo ou instruções).

Sugestão de estrutura do repositório (não obrigatória):

```
/  
  data/  
    raw/          (CSV recebidos)  
    processed/   (dados tratados)  
  notebooks/   (explorações e análises)  
  src/          (scripts Python/SQL)  
  dashboards/  (arquivo do dashboard ou instruções)  
  ppt/          (apresentação)  
  README.md
```

7. Prazos e dinâmica

Cronograma do processo:

- Até 25/12/2025: responder o e-mail com o link do repositório Git para avaliação.
- Entre 26/12/2025 e 29/12/2025: apresentação do que você fez (PPT) para o time (você receberá por e-mail a data e o horário da sua apresentação, que será de 30min junto a diretoria do Banco Shield).
- 29/12/2025: divulgação do resultado.

8. Critérios de avaliação (o que observaremos)

- Raciocínio analítico e clareza na definição do problema (hipóteses, métricas e método).
- Qualidade técnica (Python/SQL), reproduzibilidade e organização do projeto.
- Qualidade e governança dos dados (checagens, tratamento de exceções, proposta de controles e SLAs).
- Data viz e storytelling (dashboards e apresentação com narrativa orientada a decisão).
- Comunicação e soft skills (capacidade de explicar escolhas, priorizar, estimar esforço, e propor próximos passos).
- Atitude de dono (foco em valor para o negócio, recomendações práticas e consciência de risco).

Boa sorte. O Banco Shield conta com você para impedir que a Hidra domine o mercado.