

# Teste (QI)<sup>3</sup>

## *Teste para a vaga de Estágio em Engenharia de Dados*

**Por Magno Macedo de Oliveira**

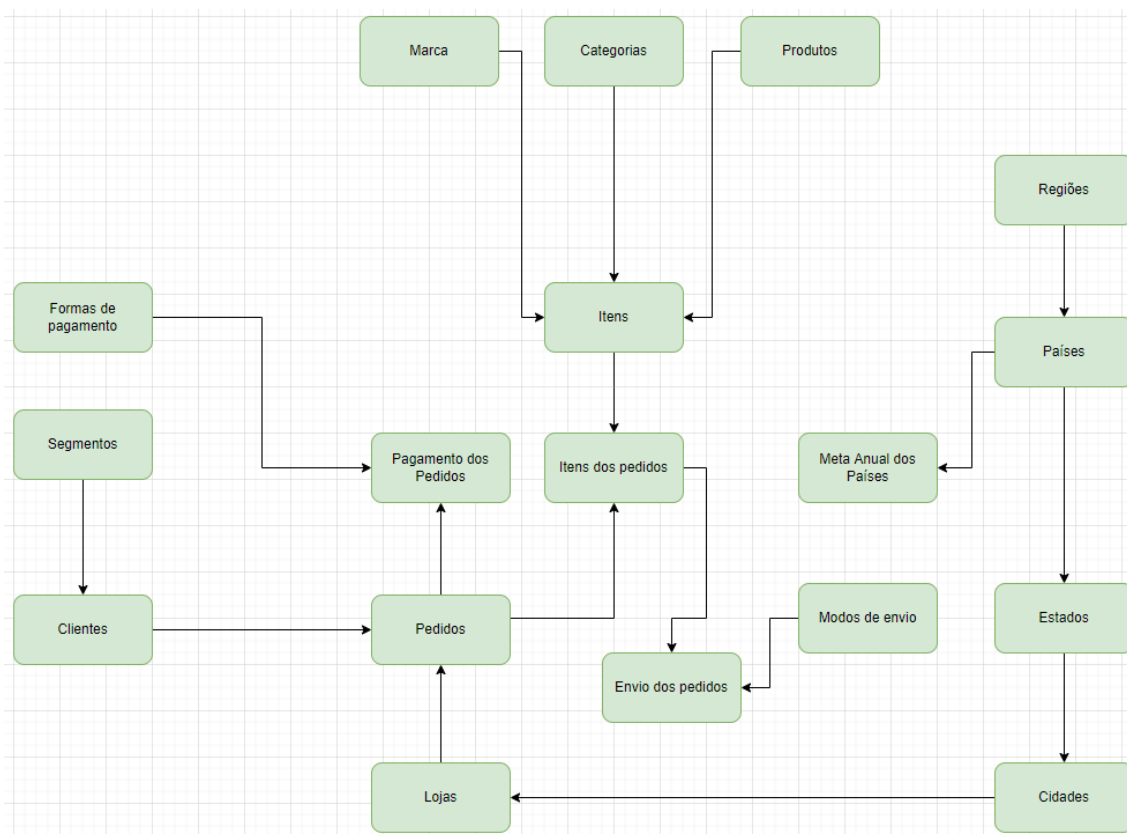
Esse projeto tem como objetivo criar uma solução voltada para tratamento, análise e visualização de dados, tendo como ponto de partida uma planilha onde consta várias abas contendo dados que simulam a de uma empresa logista internacional, indo de uma lista de marcas vendidas, cadastros de cliente e pedidos realizados em determinado período.

Para o desenvolvimento desse projeto foi utilizado:

- Google Sheets (Usado para visualizar a base original)
- Pentaho Data Integration (Ferramenta ETL)
- PostgreSQL (SGBD usado para criar o banco que recebeu os dados tratados)
- pgAdmin (Interface usada para acompanhar todo processo de criação do banco e também para realização de algumas consultas)
- DBeaver (Foi útil para geração do Diagrama de Entidade-Relacionamento do banco criado)
- PowerBI (Usado para a etapa de visualização dos dados)

## **Como foi feito?**

Com o teste em mãos ([Clique aqui](#)), meu primeiro desafio foi o de verificar os dados e começar a trabalhar em como iria fazer todo processo de ETL (Extract, Transforma and Load) para então conseguir gerar as informações solicitadas.



Esboço criado para entender melhor a base original.

A do esboço acima foi bem útil para decidir quais seriam as primeiras tabelas do banco, além de facilitar a identificação de onde existiria mais dificuldade.

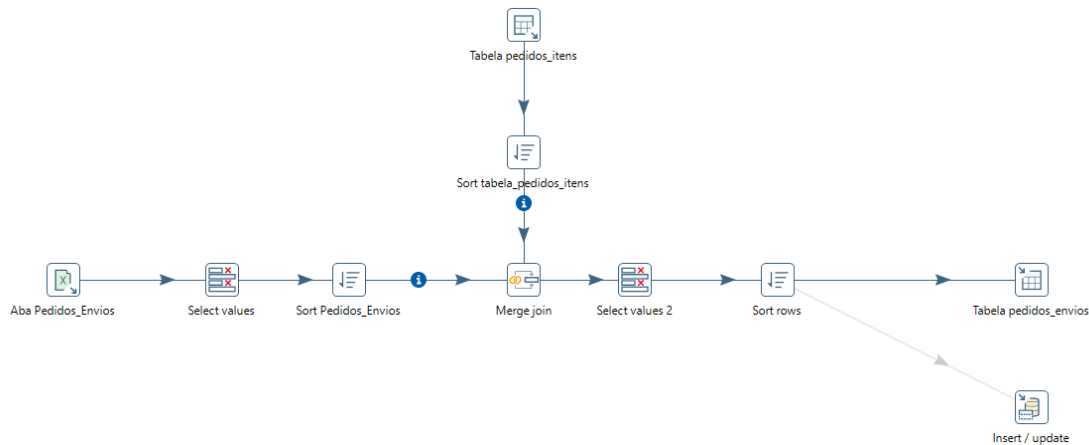
Após isso, com o PDI (Pentaho Data Integration) comecei a focar nas informações mais genéricas que não dependiam de nenhuma informação externa e então passei a criar as primeiras transformações.



*Transformação criada para abastecer a tabela Segmentos.*

As tabelas iniciais basicamente seguiram o mesmo molde, a extração dos dados da planilha base, pequeno tratamento no tipo dos dados e criação e abastecimento dos dados no banco de dados.

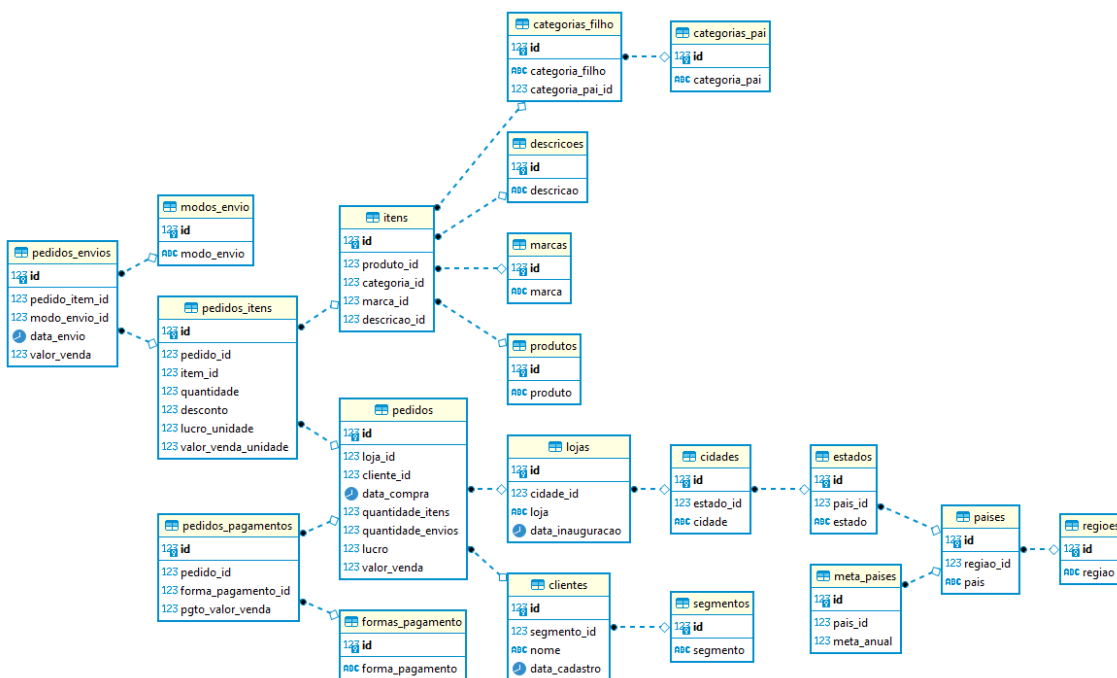
Por fim, ficaram as transformações mais elaboradas.



*Transformação criada para abastecer a tabela pedidos\_envio.*

As informações que dependiam de dados externos, dados das demais abas da planilha, precisaram de um pouco mais de cuidado, por conta de ter dados nulos, datas com formatos incorretos, linhas duplicadas e afins.

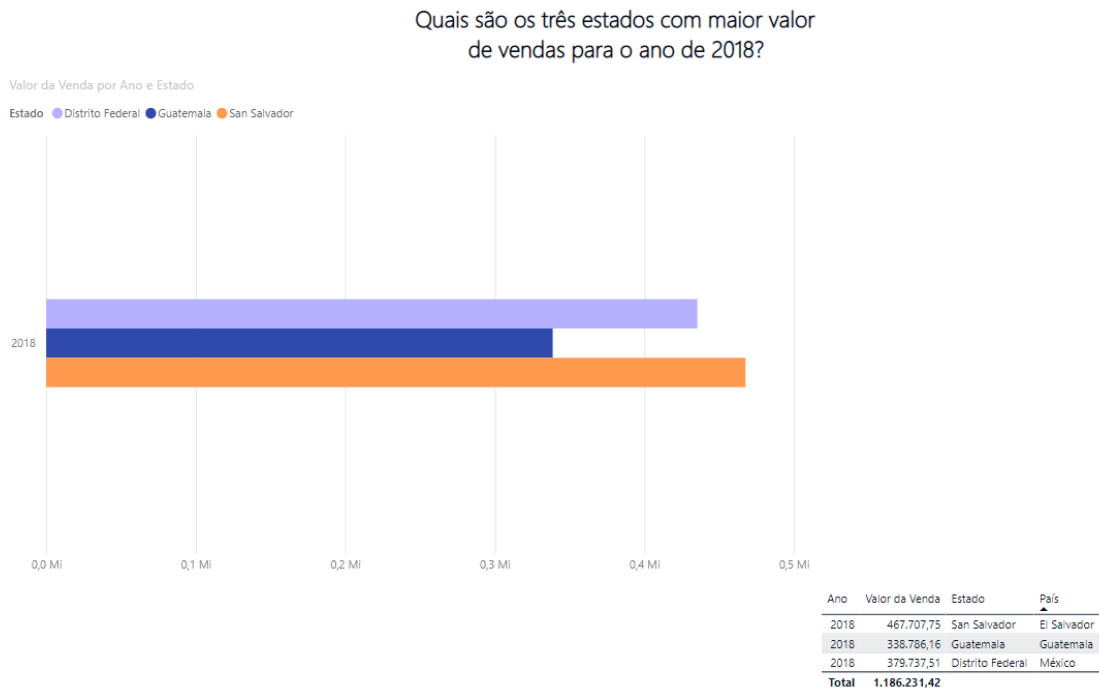
Concluído esse processo, tive como resultado a criação do banco de dados finalizada.



Banco de dados criado e abastecido.

Por fim foi realizado a criação das consultas que para atender os questionários existentes no teste.

Como eu tive um pouco de dificuldade, eu optei por algumas questões tentar fazer e apresentar diretamente no pgAdmin, salvando a Query feita além de prints da consulta junto ao resultado.



Exemplo de visualização feita com o PowerBI

Query Editor   Query History

```
1 /* Questão 05 - Quantos pedidos continham os itens 512 e 1675 na mesma compra? */
2
3 select pedidos.id as Pedido_ID, array_to_string(array_agg(itens.id), ',') as Itens_Pedido from pedidos
4 inner join pedidos_itens
5     on pedidos.id = pedidos_itens.pedido_id
6 inner join itens
7     on pedidos_itens.item_id = itens.id
8 where itens.id in(512,1675)
9 group by pedidos.id
10 having count(itens.id) >=2
```

Data Output   Explain   Messages   Notifications

	pedido_id integer	itens_pedido text
1	2415	512,1675
2	4382	512,1675

*Exemplo de resolução feita diretamente no pgAdmin*

## **Dificuldades do projeto**

No decorrer do teste me deparei com alguns desafios, porém o que considerei maior foi a de conseguir fazer uma apresentação dos dados. Me faltou um pouco de experiência com o PowerBI e quando pensei em ir para o Excell já não tinha tempo hábil. Entretanto ter realizado as queries foi uma forma de mostrar minha capacidade de ir atrás da informação e espero que isso conte na avaliação.

Agradeço a oportunidade.

**Fim.**