

Exercício: Mesclando Duas Fontes de Dados

Introdução

Você deve entender o conceito de mesclagem de dados e suas aplicações práticas. Durante esta semana você aprendeu sobre os conceitos de tipos de join, o que é join key e onde pode ser usado. Lembre-se: joins são uma das operações mais básicas e importantes no Microsoft Power BI e geralmente em todos os sistemas de banco de dados.

Neste exercício, você terá a oportunidade de aplicar seu conhecimento usando o Power Query para juntar duas fontes de dados, explorar os relacionamentos entre duas tabelas e combinar os dados relacionados.

Caso de Estudo

A Adventure Works possui dados com detalhes como Sales Order ID, Order Date, Product Key, Quantity, Unit Price, Reseller Key, Product, Reseller e alguns outros campos armazenados em tabelas separadas. Essas tabelas têm relacionamento com a tabela Sales por causa da arquitetura de dados relacional.

Como você aprendeu antes, embora essa estrutura de dados relacional forneça integridade de dados, ela mostra apenas os valores-chave. Para listar os outros detalhes das tabelas relacionadas, você deve mesclar tabelas com joins.

Seu gerente, Adio Quinn, atribuiu a você a tarefa de listar os dados de Sales com as colunas Sales Order ID, Order Date, Product Name, Quantity e Unit Price. Para fazer isso, você deve mesclar as tabelas Sales e Product criando uma query no Power BI que mescle os dados.

Adio envia dois arquivos .xlsx contendo os dados de vendas da Adventure Works chamado Sales.xlsx e dados de produtos chamado Product.xlsx. Os datasets têm uma coluna comum chamada ProductKey. Você usará a coluna comum para combinar as colunas. Você precisa ler dados das duas tabelas.

Objetivos

- Entender como combinar dados através da mesclagem de dados
- Aprender a mesclar dados no Power Query
- Explorar relacionamentos entre duas tabelas e combinar dados relacionados

Instruções

Crie um novo projeto Power BI chamado Exercise – Merging two tables. Siga as etapas abaixo para completar o exercício.

Passo 1: Baixar os Arquivos Excel

Baixar os arquivos Sales.xlsx e Product.xlsx que serão usados neste exercício.

Passo 2: Criar Projeto Power BI

Criar um projeto Power BI e abrir o Power Query Editor. Criar um novo projeto Power BI chamado Exercise – Merging two tables.

Passo 3: Abrir Power Query Editor

Abrir o Power Query Editor e importar seus datasets: Sales e Product.

Passo 4: Mesclar Queries

Após selecionar os dados de Sales no painel Queries, selecionar Merge Queries.

Na janela aberta:

- A tabela Sales será mostrada automaticamente na seção superior da janela
- Escolher a próxima tabela para mesclagem: Product
- ProductKey é a coluna comum entre as tabelas, então clicar nas colunas ProductKey em cada tabela
- No dropdown Join Kind, escolher o tipo de join Left Outer Join, que seleciona todos os registros da tabela esquerda e registros correspondentes da tabela direita

Passo 5: Selecionar Coluna(s) de Product

Após mesclar as tabelas, uma nova coluna chamada Product é adicionada ao lado direito dos dados de Sales. Isso permite escolher colunas da tabela Product.

Selecionar a coluna chamada Product da tabela Product.

Passo 6: Escolher e Reordenar Colunas de Sales

Após adicionar a nova coluna Product, ela é adicionada à query Sales como Product.Product. Você deve renomear esta coluna como Product para evitar confusão.

- Mover o campo Product da direita para a esquerda
- Remover as colunas indesejadas: Product Key (nome do produto foi adicionado pelo merge, então você não precisará do valor-chave do produto), Reseller Key, Employee Key e Sales Territory Key

- Reordenar a lista final conforme indicado na tarefa: Sales Order Number, Order Date, Product, Quantity e UnitPrice