Activity 3: Combining data sources

Learner Guide

Contents

How to use this workbook

Activity
Alongside this icon you'll find details of the group/individual activity or a point for everyone to discuss.

Useful tool
This icon indicates a technique that will help you put what you learn into practice.

Important idea or concept
Generally, this icon is used to draw your attention to ideas that you need to understand by this point in the course. Let your trainer know if you do not understand or see the relevance of this idea or concept.

Helpful hint
This icon guides you to tips or hints that will help you avoid the common pitfalls or to show you how to increase your effectiveness or efficiency in practising what you have learnt.

Key point
This icon is used to indicate something that practitioners in this field should know. It's likely to be one of the major things to remember from the course, so check you do understand these key points.

Reference material
When we have only touched briefly on a topic, this icon highlights where to look for additional information on the subject. It may also be used to draw your attention to International or National Standards or Web addresses that have interesting collections of information.

Definition
Where a word with a very specific definition (or one that could be described as jargon) is introduced, this will highlight that a definition is provided.

Warning
This icon is used to point out important information that may affect you and your use of the product or service in question.

Introduction

In this activity, you'll apply append and merge features of Power Query Editor in PowerBI to combine data sources.

Task 1: Append data using Append Queries

> **Guided activity:**
>
> **Append data tables using query**

Scenario

Annual sales data has been exported from the CRM system into four files – one for each quarter. The data needs to be combined so that it can be reported against for the whole year. The data will be loaded into and combined in a Power BI Desktop file for further analysis.

Activity

1. Start Power BI Desktop and load the data from the Tables in the following files:

   2022_Q1_Sales.xlsx

   2022_Q2_Sales.xlsx

   2022_Q3_Sales.xlsx

   2022_Q4_Sales.xlsx

   Select Get data> Excel workbook> and find the files in your browser window. All tables should be loaded individually.

2. On the last table upload, click **Transform data** to start the Power Query Editor. Alternatively, click transform data on the home ribbon.

3. Right-click on the **tblQ1Sales** table in the Query Editor and select **Reference**. This will create a refreshable snapshot of the query and leave the original intact.

4. Rename the new query as **All Sales** under Query Settings and Properties on the right-hand side of the Editor.
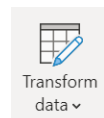
5. Ensure the All Sales query is selected. Then, use **Append Queries**, selecting **Three or more tables** to append the **tblQ2Sales** data by clicking the Add button.

6. Repeat to add **tblQ3Sales** and **tblQ4Sales** data. Click **OK**.

7. Click the **Close & Apply** button.

8. How many records are in the combined table? (Should be 2155).

9. Save the pbix file as **12 Month Sales Report.pbix** and leave it open.

Task 2: Combine data using Merge Queries

Guided activity:

Merging data sources using Query Editor

Scenario

Two separate CSV files have been created from different data sources. One contains details of our customers and the other holds product information. They need to be added to the 12 Month Sales Report. The Query Editor will be used to edit, load, and then merge the data.

Activity

1. Ensure the **12 Month Sales Report.pbix** is open in the Query Editor.

2. Stay in the Query Editor and select **New Source** to load **2022 Products.csv** into Power BI Desktop. This time select **Text/CSV** from the drop-down menu.

3. Repeat the action using **New Source** to load the **2022 Customers.csv** data into Power BI Desktop.

4. Notice that the headers have not been identified by Power BI Desktop. This is most likely due to the data in each column being text and Power BI cannot distinguish the headers from the data. Add a step to the query with the **Use First Row as Headers** command.

5. The sales report will need to provide the capability for the business to drill down from country to city and then customer. This means that

the **All Sales** query will need to have a reference to the **City** column from the **2022 Customers** query.

6. Ensure that the **All Sales** query is selected on the left and then click **Merge Queries** on the Home tab.

7. In the Merge dialog box, select the **2022 Customers** table in the dropdown menu and then select the **Company Name** in both tables as the matching columns.

8. How many rows are matched?

9. Click **OK**, followed by the **Expand** button on the **2022 Customers** header.

10. In the dialog box, uncheck the **Select All** option and then select **City**. Click **OK**. Leave the column name as it is. This will make it easier to identify the source of the data.

11. Click **Apply** to push the changes through to the front of Power BI Desktop and leave the Query Editor open.

Task 3: Append and merge data

> **Independent activity:**
>
> **Append and merge data from a workbook**

Scenario

An additional reporting requirement has been identified in the 12 Month Sales Report. The products included in the sales need to be grouped by category such as Beverages or Dairy Products. The category names are stored in a table in an Excel workbook and can be linked to the Products csv file by Category ID.

Activity

The categories will need bringing into the 12 Month Sales Report and then the Category must be added to the All Sales table.

1. Use the Query Editor to load the T_Categories table from the BI_NWind_All.xlsx workbook.

2. Remove the Picture column as it contains an image that is not required.

3. Use Merge Queries to display the Category Name column in the All Sales table. It should match all 2155 records.

4. When you have completed the merge, Close & Apply the changes. Save changes to the 12 Month Sales Report file. Close it when saved.