**About the final project and assessment: Extract, transform and load data in Power BI**

**Introduction**

As you near the end of the *Extract, Transform, and Load Data in Power BI* course, you will undertake a comprehensive final quiz that aims to assess your comprehension and proficiency in the course content.

This evaluation will not only gauge your mastery of the learning objectives but also identify areas that may require additional focus for a thorough course completion. Ultimately, the graded course assessment aims to validate your ability to apply the acquired knowledge to the quiz questions and real-world scenarios.

**What to expect from the graded course assessment**

The graded course assessment consists of 30 graded quiz questions based exclusively on the topics you’ve covered throughout the course, so there shouldn’t be any unexpected elements in the assessment. It will take about 1 hour and 30 minutes to complete, and you need to score 80% to pass.

You can take the quiz more than once if you don’t achieve this score but note that the questions will vary with each attempt.  As you progress with the questions, review the feedback on your answers. Where necessary, revisit and work through the topics that you feel require your further attention.

**How to prepare for the graded course assessment**

This assessment will test you on the topics covered throughout this course, including:

* Data Sources supported by Power BI
* Local datasets vs. shared datasets
* Storage modes in Power BI
* Structured vs. unstructured data
* PowerBI Connectors
* Triggers and Actions
* Data Types in Power BI
* The Applied Steps List
* Common Data Errors
* Pivot and Unpivot Columns
* Combining Tables with append
* Combining tables with merge
* Join Types
* Join Keys
* Loading Data
* The Staging Area in Power BI
* Profiling Data in Power BI
* Using the Data Profiling Tools
* Reference queries and dataflows
* Query Parameters
* Global options for files

**Final project assessment**

In this project assessment, you will have the opportunity to apply your knowledge in an end-to-end scenario by using Power Query to clean and transform multiple data sources and join and merge them. This project assessment will allow you to test yourself on the various skills you learned in this course, as outlined above.

**Power BI Desktop User Interface Disclaimer**

Power BI Desktop is updated and released on a monthly basis, incorporating customer feedback and new features. You might experience changes in the Power BI Desktop User Interface (UI) that have taken place after the development of this training content. As a result, the screenshots in the videos, readings, or exercises might not align exactly with how you experience the UI. However, please note that these changes do not impact the functionalities of the UI. Hence, you will still be able to perform all the steps shown in that video, reading, or exercise.

**Conclusion**

For optimal readiness, it is recommended to revisit the quizzes you have previously completed in this course, paying special attention to any incorrect answers. Take advantage of the feedback provided, using it as a roadmap to prioritize the topics that require further attention. Additionally, revisiting relevant videos, readings, and exercises from the course can prove beneficial in enhancing your mastery of the content.

**Final Course Project: Transforming multiple data sources**

[Order2022](https://d3c33hcgiwev3.cloudfront.net/nc1fW78PS22NF8moeFiTZA_5950e1bc92374abc8798250328b155e1_Order2022.xlsx?Expires=1709856000&Signature=DLjcifPJvcB7wCtS7XQU0x49J3FW6IgVGgT-zGVW6JMRGWXEPODkXmZysLPkOJbPn1m~2t1KIpfNdIid3EJnIoDkTTmgFar28Fxivivkd2778m6yTDDqmxA8s3LU59gGWyPyp7jZxLor90gz3TdoI9qzJJcN1Hiswrq1XF9D4qU_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[XLSX File](https://d3c33hcgiwev3.cloudfront.net/nc1fW78PS22NF8moeFiTZA_5950e1bc92374abc8798250328b155e1_Order2022.xlsx?Expires=1709856000&Signature=DLjcifPJvcB7wCtS7XQU0x49J3FW6IgVGgT-zGVW6JMRGWXEPODkXmZysLPkOJbPn1m~2t1KIpfNdIid3EJnIoDkTTmgFar28Fxivivkd2778m6yTDDqmxA8s3LU59gGWyPyp7jZxLor90gz3TdoI9qzJJcN1Hiswrq1XF9D4qU_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[Order2023](https://d3c33hcgiwev3.cloudfront.net/CwXbeUPnRumGxY8Z-yxIhA_7c3b1538581d4034b624f2e8b2d0eea1_Order2023.xlsx?Expires=1709856000&Signature=JHjNvJVIbA-LJe8-s3CkswzX10M~pDJKllMyjbGgcj2q31xd~AuXa-k893gC5JlFnu9x3sTUmpgnaf98WSNSjVpb1ZUgUIuaUgatvkTG0gC1oKBGycqCgOCwMxnnUzU2~i99DQH5v1FWSxkJuTKjscX9LlAA7Vo~6XTrmY4W3-A_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[XLSX File](https://d3c33hcgiwev3.cloudfront.net/CwXbeUPnRumGxY8Z-yxIhA_7c3b1538581d4034b624f2e8b2d0eea1_Order2023.xlsx?Expires=1709856000&Signature=JHjNvJVIbA-LJe8-s3CkswzX10M~pDJKllMyjbGgcj2q31xd~AuXa-k893gC5JlFnu9x3sTUmpgnaf98WSNSjVpb1ZUgUIuaUgatvkTG0gC1oKBGycqCgOCwMxnnUzU2~i99DQH5v1FWSxkJuTKjscX9LlAA7Vo~6XTrmY4W3-A_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[OrderDetails](https://d3c33hcgiwev3.cloudfront.net/Jh-UQm8yTX2goVaLkDZxvg_be4a12ff4a3d42ebaaaa9518486abfa1_OrderDetails.xlsx?Expires=1709856000&Signature=SQcaCpujObVLaTDwmDqfIbLcBFPVwAt0DhCfTfPdadmNxd4XC4PlKdM1gGfI3L0XVk6cyazTU6YBOqBa1QBBNtAPzjCE96ot7lhU9Ys1G7RmUf-UOPMVHEGdx2GG9K8V-NcGxn1RxAA4yU7Wv-5d1fcVeknxHwPVMeVoUcYOXro_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[XLSX File](https://d3c33hcgiwev3.cloudfront.net/Jh-UQm8yTX2goVaLkDZxvg_be4a12ff4a3d42ebaaaa9518486abfa1_OrderDetails.xlsx?Expires=1709856000&Signature=SQcaCpujObVLaTDwmDqfIbLcBFPVwAt0DhCfTfPdadmNxd4XC4PlKdM1gGfI3L0XVk6cyazTU6YBOqBa1QBBNtAPzjCE96ot7lhU9Ys1G7RmUf-UOPMVHEGdx2GG9K8V-NcGxn1RxAA4yU7Wv-5d1fcVeknxHwPVMeVoUcYOXro_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

**Introduction**

In this course, using Power BI, you covered the data extract, transform, and load process (ETL). You should now have a good understanding of:

* Connecting to data sources
* Column data types
* Common data errors
* Combining tables with merge and append
* Advanced data transformation techniques
* Loading and staging data
* Profiling data by using column quality
* Column distribution and column profile, and its practical application in identifying data anomalies.

In this exercise, you will apply your knowledge in an end-to-end scenario by using Power Query to clean and transform multiple data sources and join and merge them. You’ll also examine the valid, error, empty, min, max, unique, and distinct values in the rows. This will allow you to identify the anomalies in the data. Finally, you will remove the data sources with anomalies. This exercise will help you understand how to clean, transform, join, and merge data sources in Power Query, and identify potential data anomalies by using data profiling tools.

**Case study**

You are working as a data analyst at Adventure Works. Sales data is contained in two main tables, Order and OrderDetails.

**Data Sources**

The **Order** data table includes general information about the sales such as:

| **Order** |  |  |  |
| --- | --- | --- | --- |
| OrderDate | TotalDue | TerritoryID | SalesPersonID |

The **OrderDetails** data contains related details of each main sales record such as:

| **OrderDetails** |  |  |  |
| --- | --- | --- | --- |
| ProductID | OrderQty | UnitPrice | UnitPrice Discount |

**Files**

Adventure Works trades internationally and generates a large volume of sales data. To manage file sizes, the active Order table only includes data for the year 2023. Older data is stored in separate files for each year with the same fields and table structure.

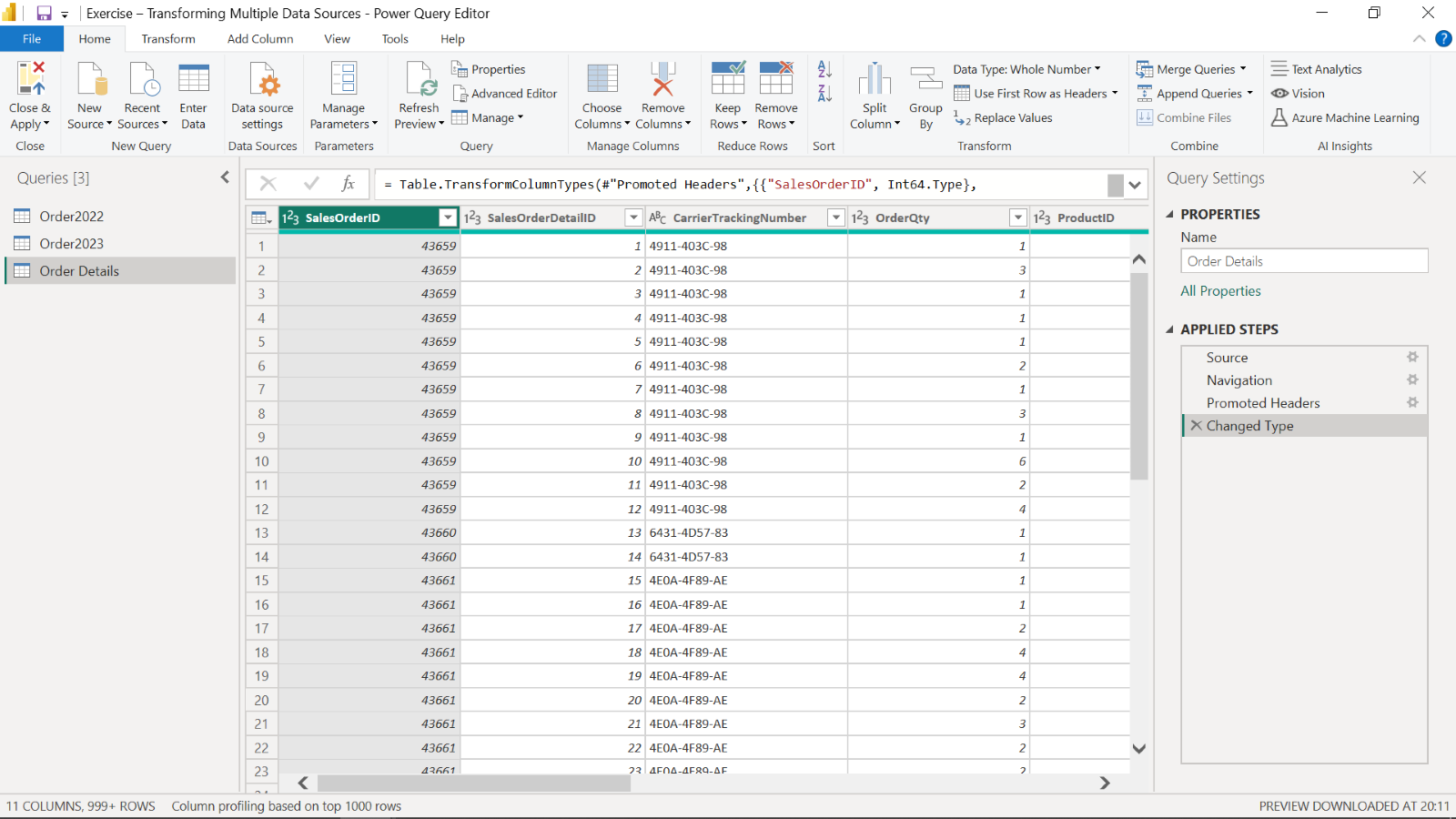
**The task**

Your manager, Adio Quinn, asks you to conduct a detailed analysis of store sales. In the detail table, **OrderDetails**, there are multiple fields, but you only need **ProductID**, the quantity sold (which is in the field **OrderQty**), and the **UnitPrice**. Therefore, you are expected to remove unnecessary fields, and also eliminate empty rows, and identify any anomalies to remove those rows if necessary. After performing these tasks, you will append the two separate sales data sources together and then merge that with **OrderDetails.** Follow the steps below to complete the exercise.

**Instructions**

**Step 1: Set up the project**

1. Create a new Power BI projectcalled *Exercise – Transforming Multiple Data Sources.*
2. Download the *Order2022.xlsx*, *Order2023.xlsx* and *OrderDetails.xlsx* files, which you will use in this exercise.



**Step 2: Open the Power Query Editor**

1. Use the **Get Data** feature in Power BI.
2. Select **Transform** to open the Power Query editor.
3. Import your .xlsx datasets, *Order2022*, *Order2023* and *OrderDetails.*

**Step 3: Choose columns from Order Details**

1. Open the **Order Details** query by selecting it from the Queries pane.
2. Keep **SalesOrderID**, **ProductID**, **OrderQty**, **UnitPrice** columns.
3. Remove the other columns by simply right clicking and selecting the **Remove** option in the shortcut-menu.

**Step 4: Profile data in Order Details**

1. To profile data, select **Order Details**.
2. On the View tab, in the Data Preview group, check Column Distribution, Column Quality and Column Profile checkboxes.
3. Note the amount of **distinct** values and **unique** values in **Column Distribution**.
4. Also check the **valid**, **error** and **empty** values of all columns in **Column Quality**.

**Step 5: Detect potential anomalies in the price**

To detect potential anomalies and assess column distribution for the price, in this step you will assess the **UnitPrice** column in **Order Details** list and find out the min, max, mean values and the distribution of the values.

1. In the **View** tab, in the **Data Preview** group, check **Column Profile** while keeping **Column Distribution** checkbox as checked.
2. Note the **min**, **max**, and **mean** values for the **UnitPrice** column and also assess the **Column Distribution**.
3. You should find that three rows are outliers in the **UnitPrice** column.
4. Consider these three rows as data anomalies (they are most probably mistypes when data was entered) . Remove them by **filtering** and **unchecking** these values to avoid confusion and incorrect calculations.

**Step 6: Append queries**

1. Append **Order2022** and **Order2023** queries in a new master table.
2. Check the newly created query, its **column names**, **row number**, and the **values** appended.
3. Make sure that the operation has been completed successfully.
4. Rename it as **Orders***.*

**Step 7: Merge queries**

1. Select the **Order Details** data in the Queries pane and choose **Merge Queries**.
2. In the opened window, the Order Details table will be automatically shown in the upper part.
3. You will be choosing Order as the next table for merging. **SalesOrderID** is the common column between the tables. To begin establishing a connection select the **SalesOrderID** column in each table.
4. For the **Join Kind** dropdown, choose the join type **Inner Join**, which selects the matching records from the left table and the right table.
5. Choose **Expand** near the newly added Orders table column and choose only the **OrderDate** column from **Orders** table in the opened window.
6. Rename the **Orders.OrderDate** column to **OrderDate** by simply double-clicking on it.

**Conclusion**

You have now successfully completed an end-to-end scenario by cleaning and transforming multiple data sources, joining and merging them, and identifying potential anomalies in the data using Power Query.

# **xemplar: Transforming multiple data sources**

[Order2022](https://d3c33hcgiwev3.cloudfront.net/nc1fW78PS22NF8moeFiTZA_5950e1bc92374abc8798250328b155e1_Order2022.xlsx?Expires=1709856000&Signature=DLjcifPJvcB7wCtS7XQU0x49J3FW6IgVGgT-zGVW6JMRGWXEPODkXmZysLPkOJbPn1m~2t1KIpfNdIid3EJnIoDkTTmgFar28Fxivivkd2778m6yTDDqmxA8s3LU59gGWyPyp7jZxLor90gz3TdoI9qzJJcN1Hiswrq1XF9D4qU_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[XLSX File](https://d3c33hcgiwev3.cloudfront.net/nc1fW78PS22NF8moeFiTZA_5950e1bc92374abc8798250328b155e1_Order2022.xlsx?Expires=1709856000&Signature=DLjcifPJvcB7wCtS7XQU0x49J3FW6IgVGgT-zGVW6JMRGWXEPODkXmZysLPkOJbPn1m~2t1KIpfNdIid3EJnIoDkTTmgFar28Fxivivkd2778m6yTDDqmxA8s3LU59gGWyPyp7jZxLor90gz3TdoI9qzJJcN1Hiswrq1XF9D4qU_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[Order2023](https://d3c33hcgiwev3.cloudfront.net/4qAUEXfvTK6mEswlvamRNg_efb7b7de223443cc8164016f8d9787e1_Order2023.xlsx?Expires=1709856000&Signature=ZHa995X4B5nLyBt-4qxg9VlM5pdq9owXqO01lrGD0Wq2hE5GY9~dy86tu0KEOZjr6WtDpM9VtxNCputymvxELcHo0Uub-QHZdDSGhESLxEahwOWTPXMyfcuH3lrD1-kJNymVxR75TlUOjQTQzviY4~b7HOjhG7xB78CU4rO0IVQ_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[XLSX File](https://d3c33hcgiwev3.cloudfront.net/4qAUEXfvTK6mEswlvamRNg_efb7b7de223443cc8164016f8d9787e1_Order2023.xlsx?Expires=1709856000&Signature=ZHa995X4B5nLyBt-4qxg9VlM5pdq9owXqO01lrGD0Wq2hE5GY9~dy86tu0KEOZjr6WtDpM9VtxNCputymvxELcHo0Uub-QHZdDSGhESLxEahwOWTPXMyfcuH3lrD1-kJNymVxR75TlUOjQTQzviY4~b7HOjhG7xB78CU4rO0IVQ_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[OrderDetails](https://d3c33hcgiwev3.cloudfront.net/kGQGkdsLTkyzrxjbpmS0_g_da0ea76d222348d0aa044b7fd3802fe1_OrderDetails.xlsx?Expires=1709856000&Signature=GwNqSpyKYZdndBUe5JWJxPfpmmmoJi1qexLEP-7nKnOKKrDT7j5dhI3M6-DeTtKD~94tuHLMOUZb~dzI8ryQ3uCEQMXN0D8NEUHY0WvYY71oTBG2724-ScOIFasQnAE7f-bF~1y31fUY51BgVHuEa2d3CZ4Z4xacYM1rjJYYZ0g_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[XLSX File](https://d3c33hcgiwev3.cloudfront.net/kGQGkdsLTkyzrxjbpmS0_g_da0ea76d222348d0aa044b7fd3802fe1_OrderDetails.xlsx?Expires=1709856000&Signature=GwNqSpyKYZdndBUe5JWJxPfpmmmoJi1qexLEP-7nKnOKKrDT7j5dhI3M6-DeTtKD~94tuHLMOUZb~dzI8ryQ3uCEQMXN0D8NEUHY0WvYY71oTBG2724-ScOIFasQnAE7f-bF~1y31fUY51BgVHuEa2d3CZ4Z4xacYM1rjJYYZ0g_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

**Introduction**

In the exercise *Transforming multiple data sources,* you put into practice your understanding of how to clean and transform multiple data sources, profile data and detect potential anomalies in Microsoft Power Query.

You were asked to complete tasks, including:

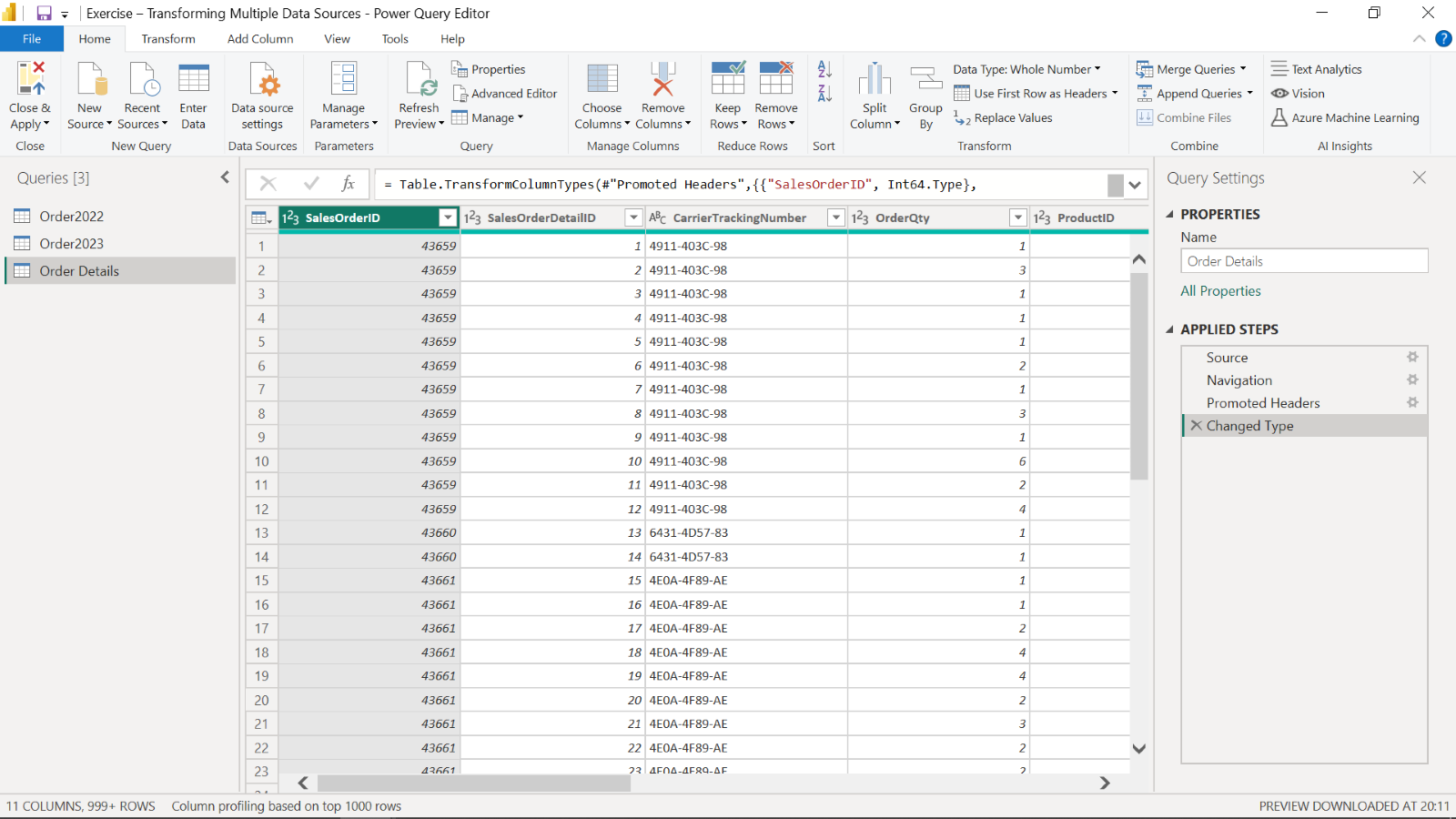
* Clean and transform multiple data sources.
* Join and merge multiple data sources.
* Examine the valid, error, empty, min, max, unique, and distinct values in the rows, to identify the anomalies in the data.
* Remove the data sources with anomalies.

This reading provides a step-by-step guide to completing these tasks, accompanied by screenshots for each step. Follow the steps below to compare the outcome with your own work on the exercise.

**Instructions**

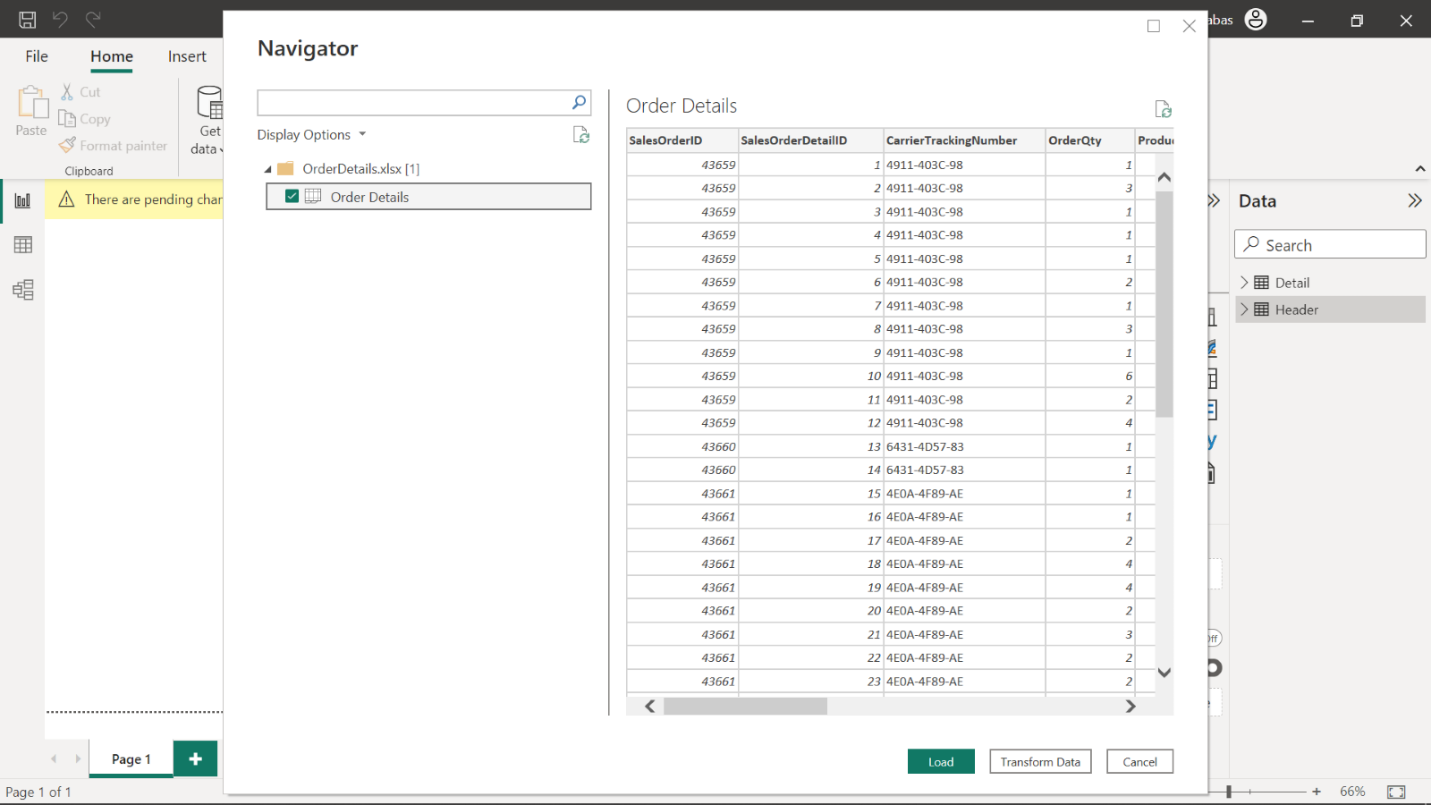
**Step 1: Setup the project**

1. You created a new Power BI project called *Exercise – Transforming Multiple Data Sources*.
2. Then you downloaded the *Order2022.xlsx, Order2023.xlsx* and *OrderDetails.xlsx* files, that were needed for this exercise.



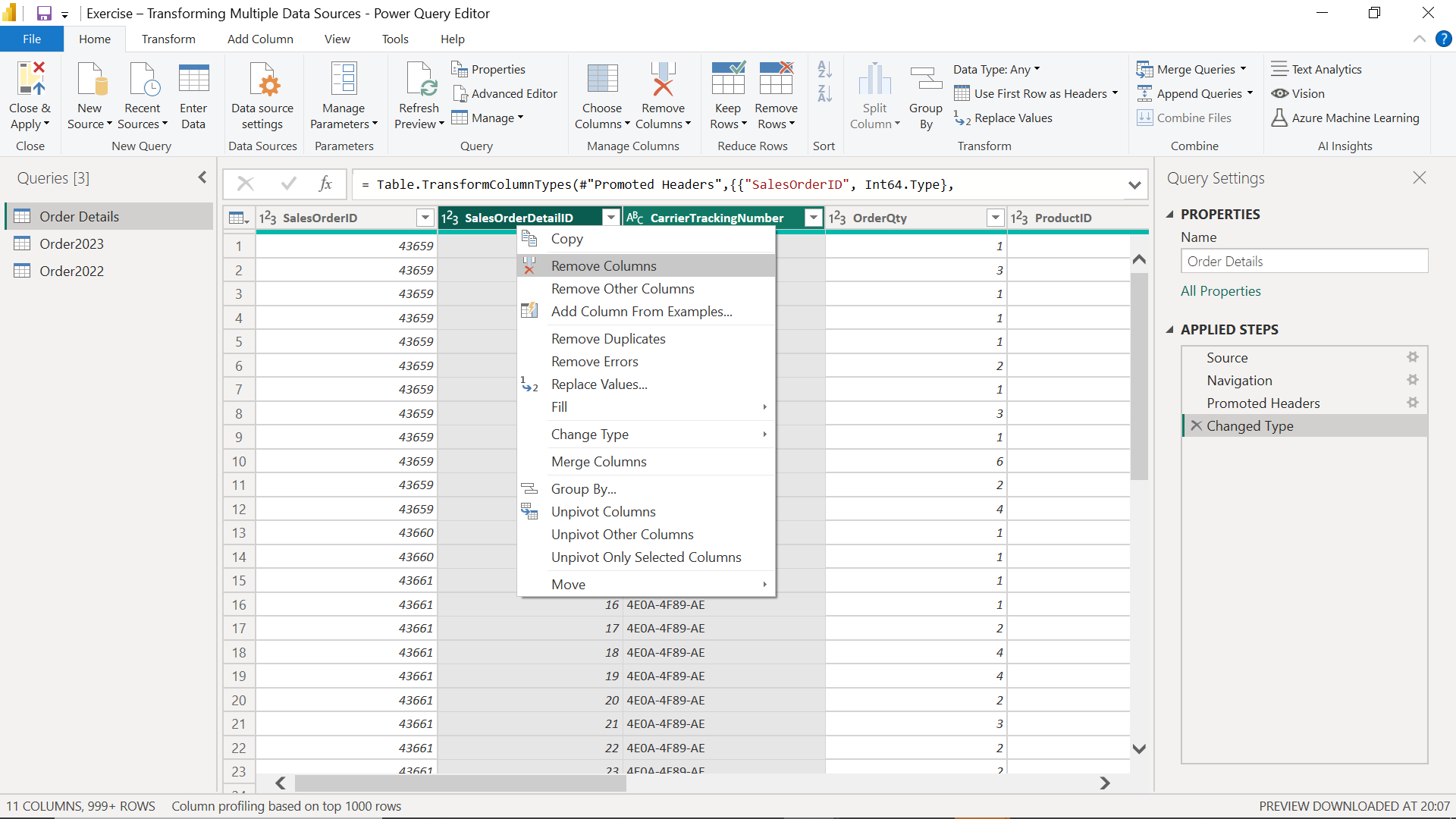
**Step 2: Open the Power Query Editor**

1. You used the **Get Data** feature in Power BI.
2. You selected **Transform** to open the Power Query editor.
3. Your *.xlsx* datasets, *Order2022*, *Order2023* and *OrderDetails* were imported. The Power Query Editor window opens. You can now begin working with the data in Power Query.



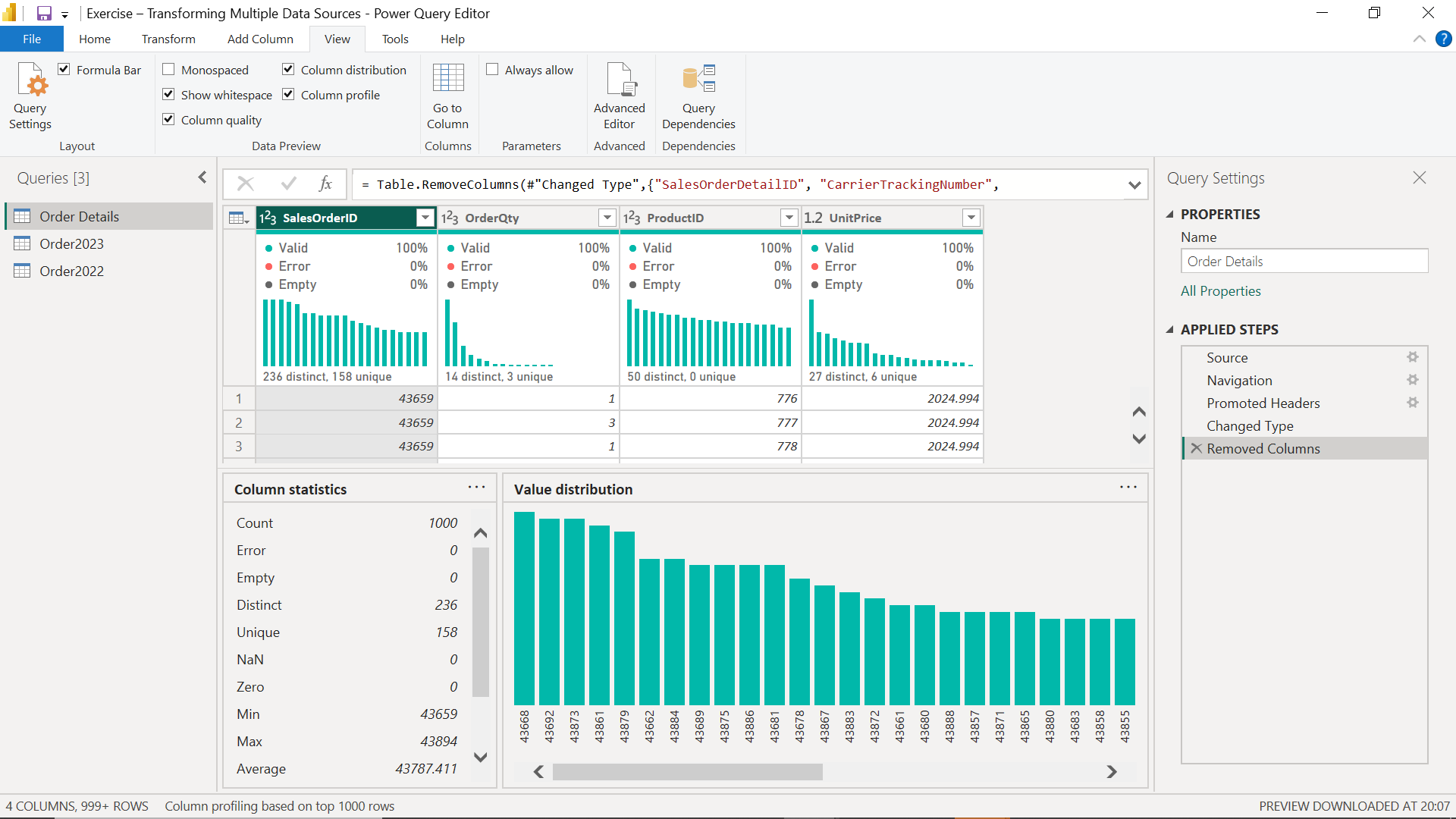
**Step 3: Choose columns from OrderDetails**

1. The **OrderDetails** query was opened by selecting it from the Queries pane.
2. You retained the columns: **SalesOrderID**, **ProductID**, **OrderQty**, **UnitPrice**
3. You removed the other columns by selecting them and using a right-click to choose the **Remove** option.



**Step 4: Profile data in OrderDetails**

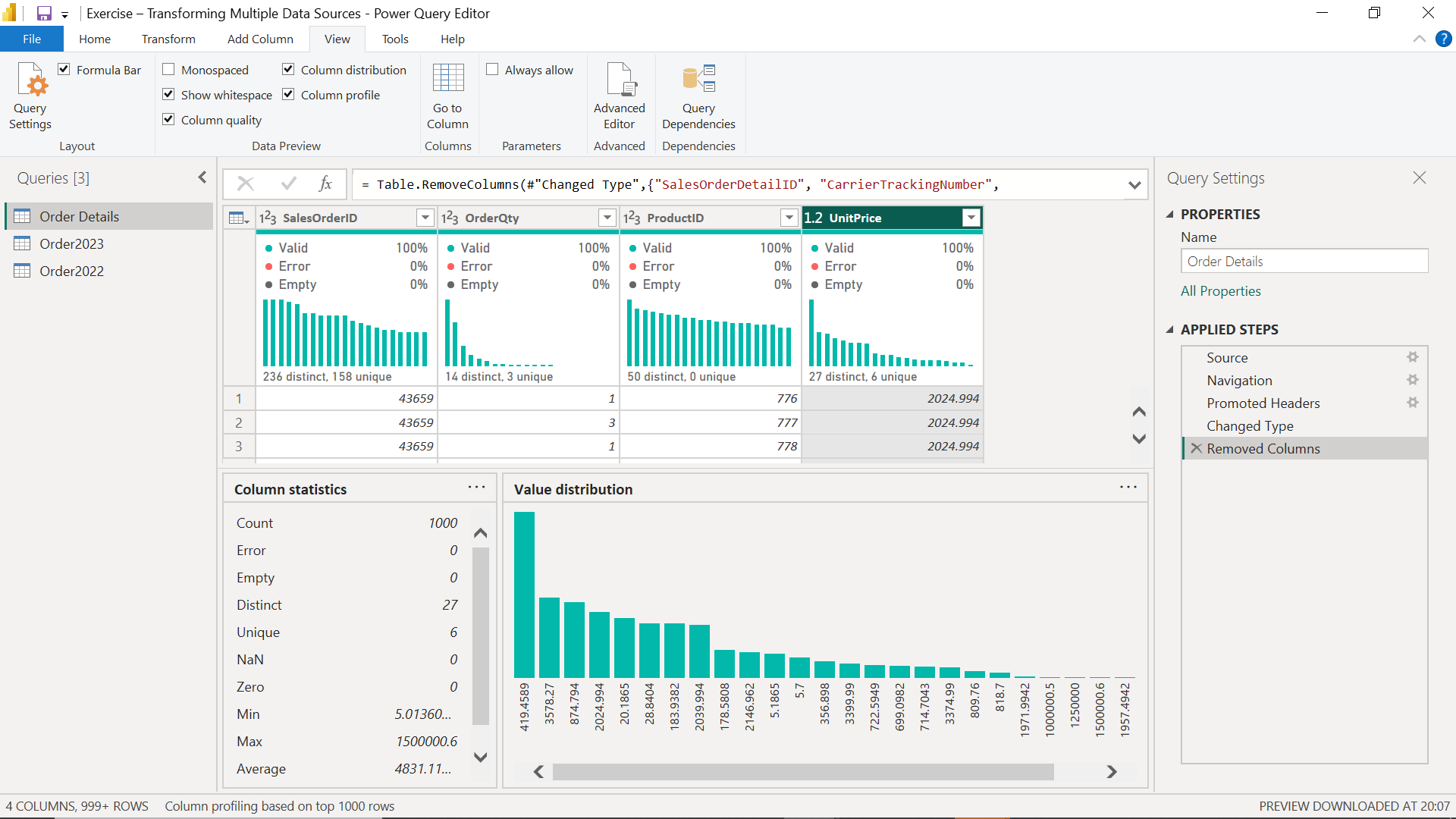
1. To profile the data, you selected **OrderDetails**
2. Then on the **View** tab, in the **Data Preview** group, you ensured there was a check on **Column Distribution, Column Quality and Column Profile** checkboxes.
3. You took note of the number of **distinct** values and **unique** values in **Column Distribution**.
4. You also checked the **valid**, **error** and **empty** values of all columns in **Column Quality**.



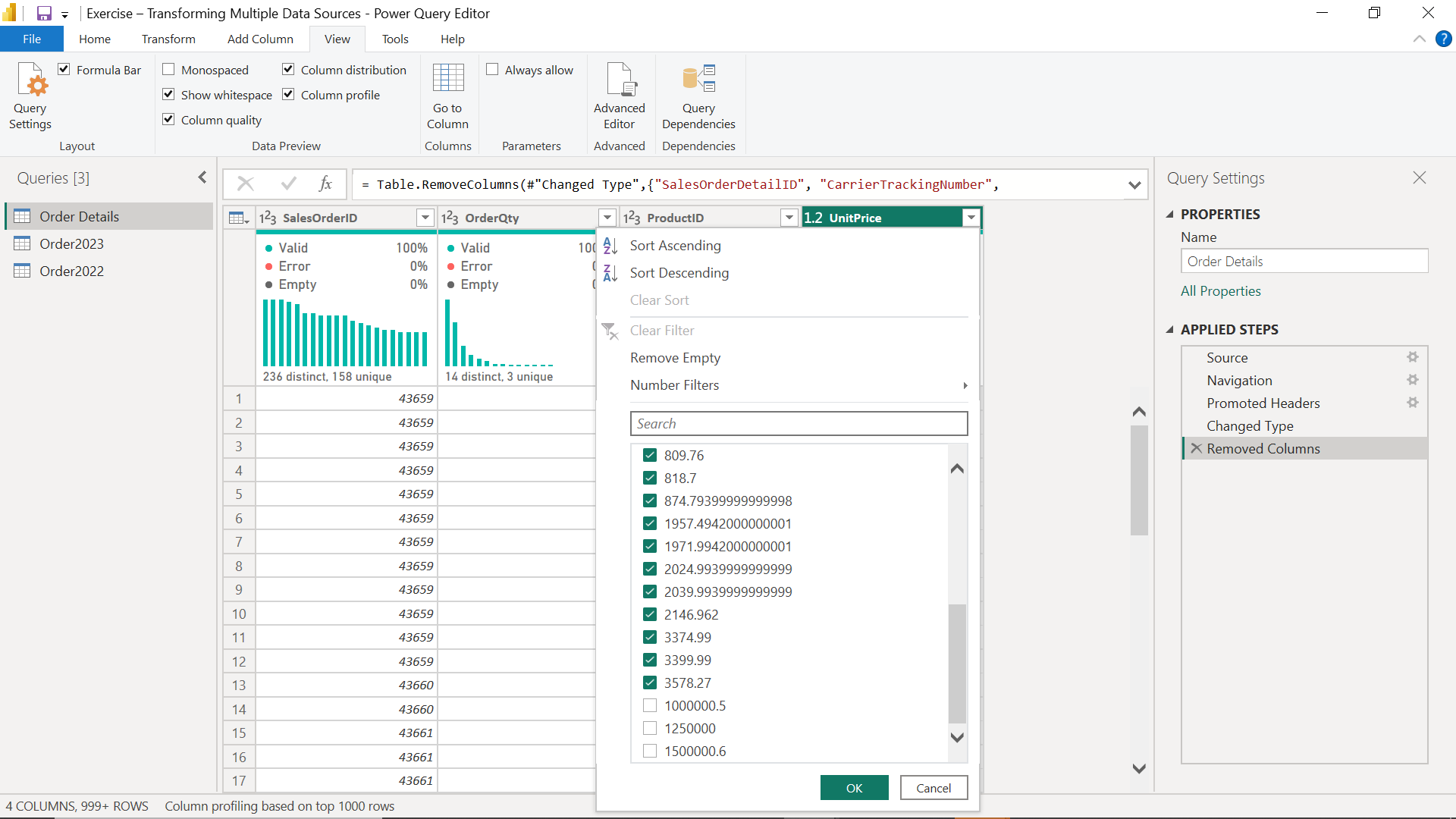
**Step 5: Detect potential anomalies in the price**

To detect potential anomalies and assess column distribution for the price, in this step you assessed the **UnitPrice** column in **OrderDetails** list and found the min, max, mean values and the distribution of the values.

1. In the **View** tab, in the **Data Preview** group, you checked **Column Profile** while keeping **Column Distribution** checkboxes as checked.



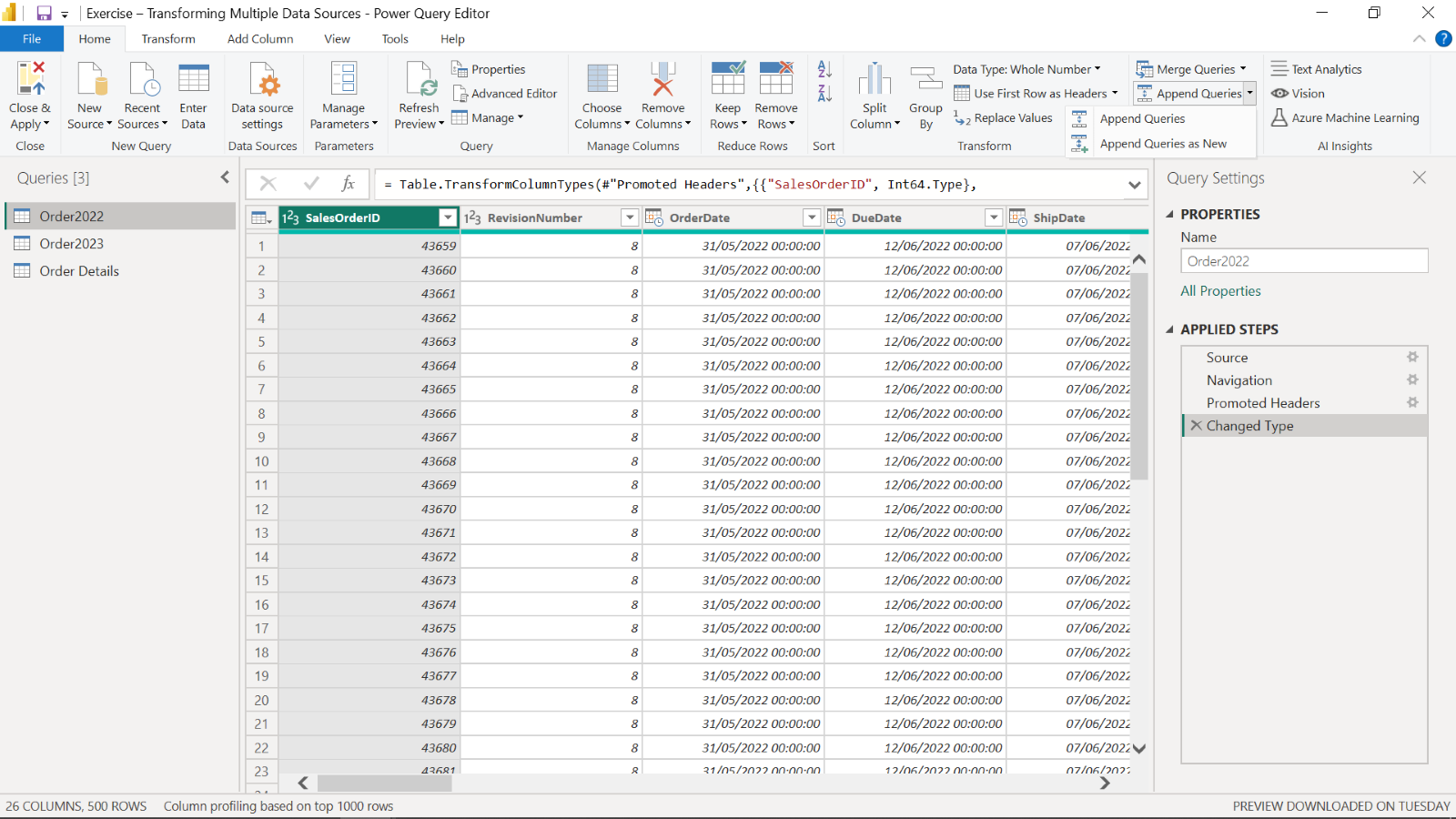
1. You noted the **min**, **max**, and **mean** values for the **UnitPrice** column and also assessed the **Column Distribution**.
2. You filtered the **UnitPrice** column by simply clicking the down arrow near the **UnitPrice** heading and find out that 3 rows are outliers in the **UnitPrice** column. You should have found that these three rows were the outliers:
3. You considered these three rows as data anomalies (they were probably mistypes when data was entered). Therefore, you removed them by **filtering** and **unchecking** these values to avoid confusion and incorrect calculations.



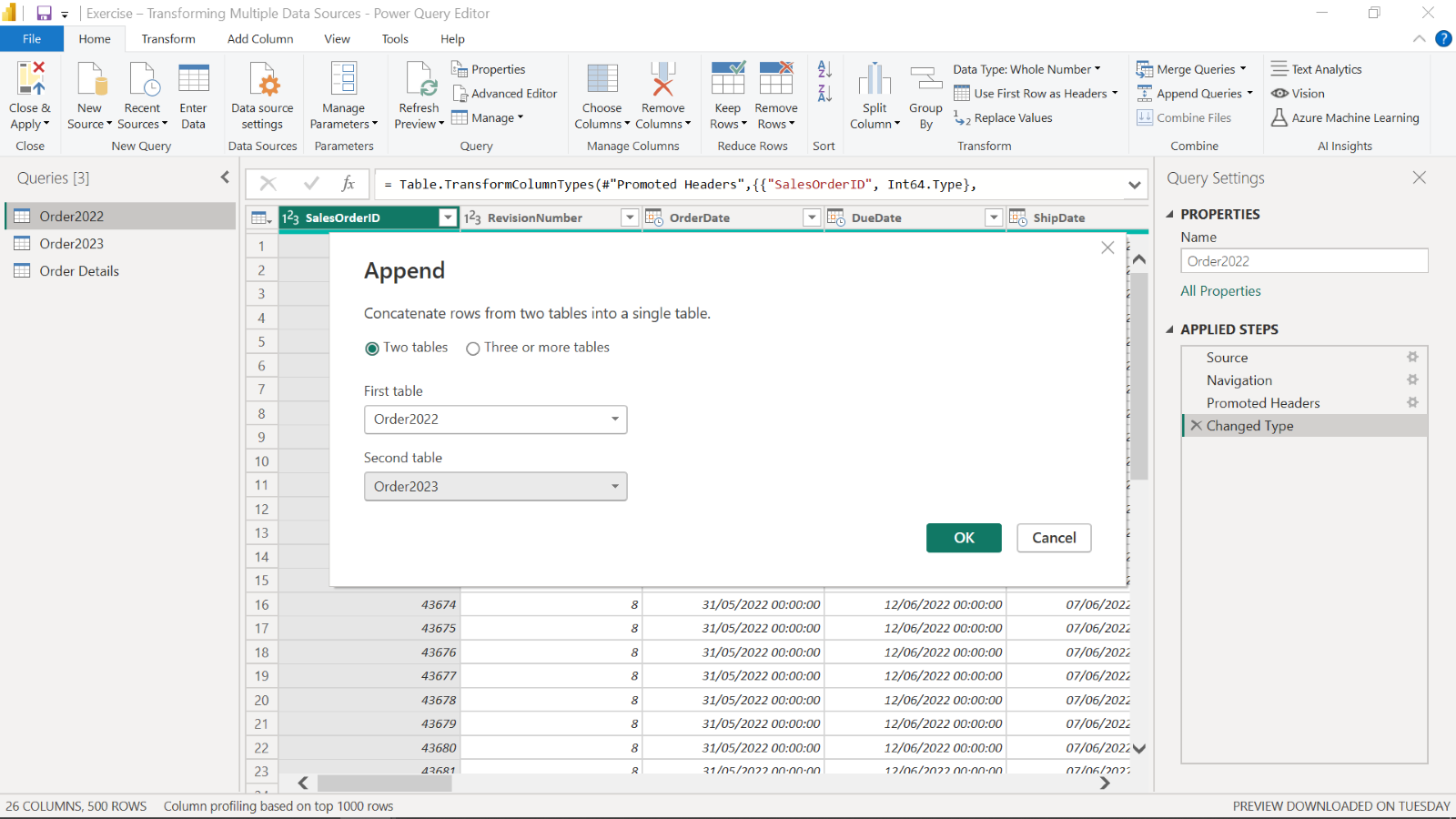
**Step 6: Append Queries**

In your task statement, you are asked to append the separate sales data together after performing the other tasks. To do this, you need to append the queries in a new query.

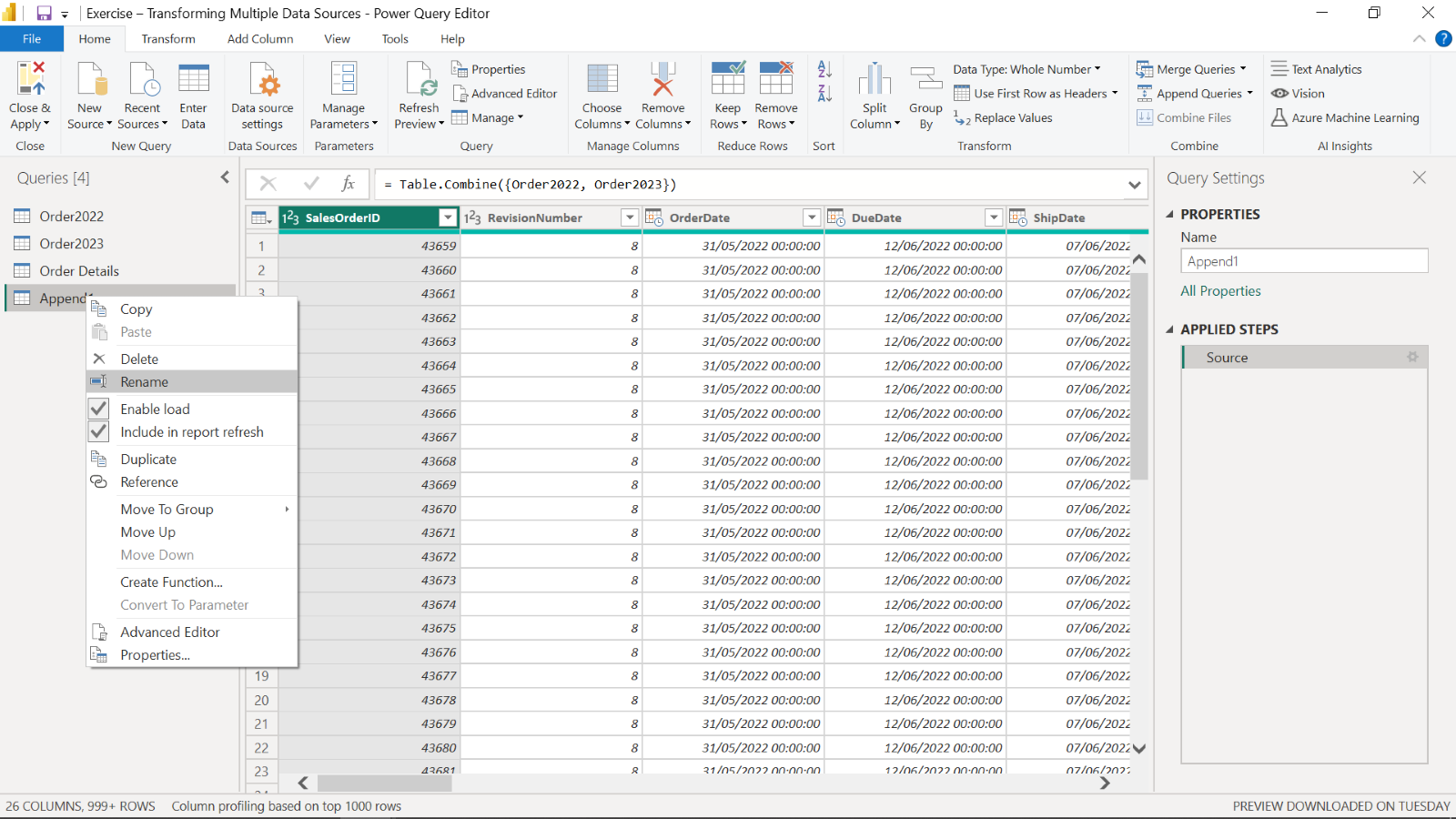
1. You select *Order2022* then on the **Home** tab in the **Combine** group you selected **Append Queries as New** option in the **Append Queries** drop-down menu.



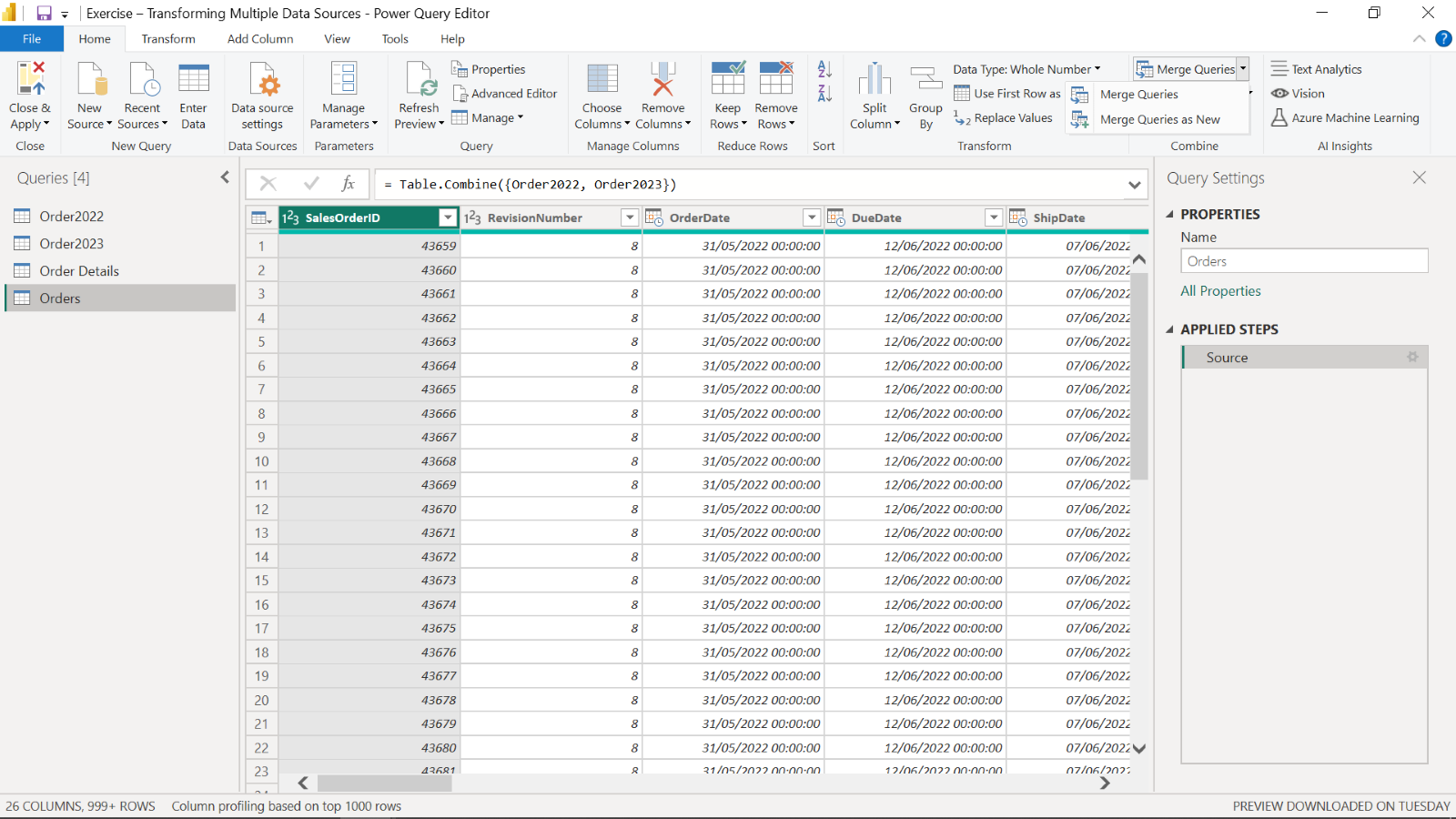
You selected *Order2022* as the first table and *Order2023* as the second table.



1. You then checked the newly created query, its **column names**, **row number**, and the **values** appended.
2. You checked to make sure that the operation has been completed successfully.



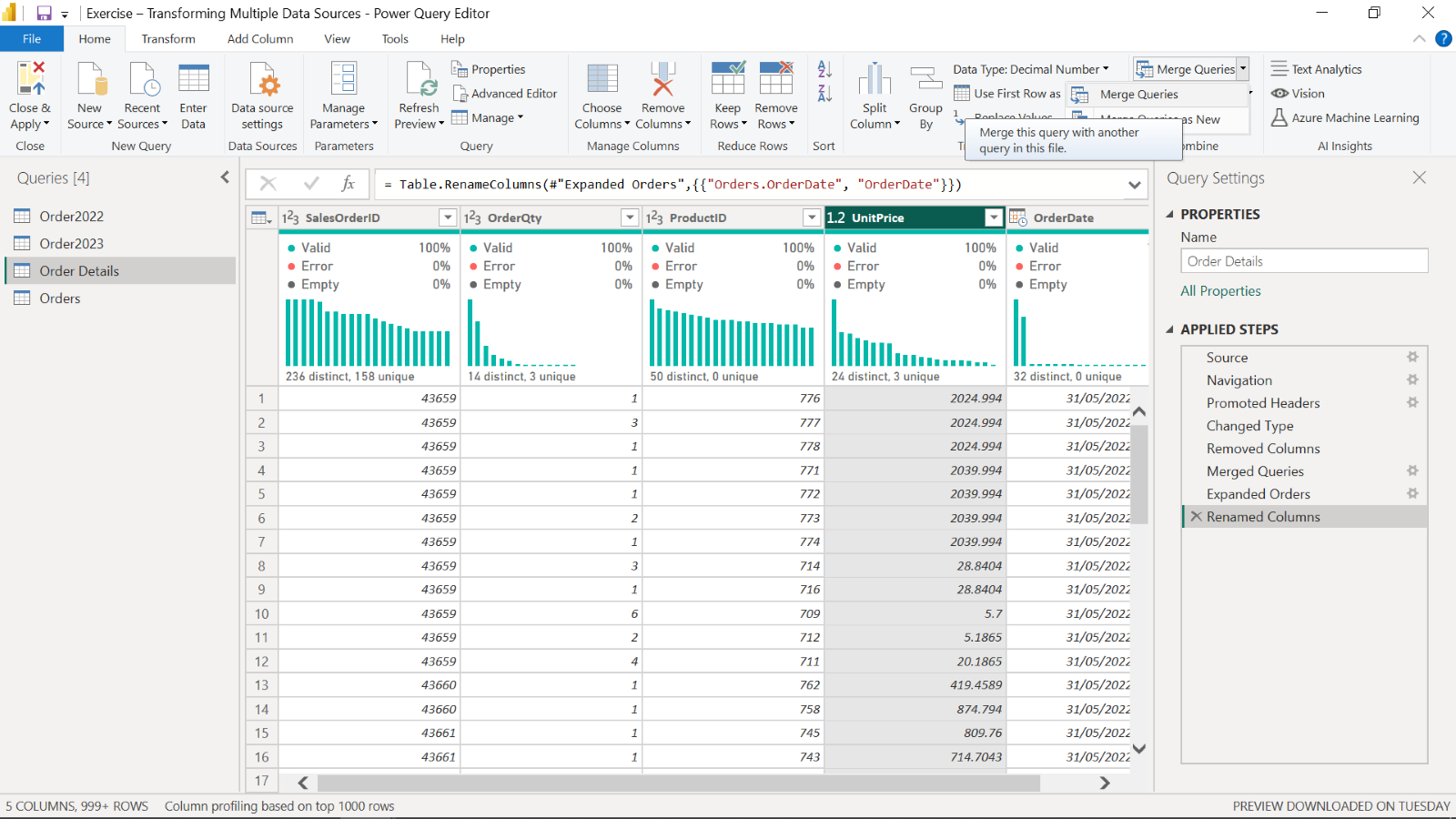
1. You right-clicked on the newly added query and selected **Rename** to rename it as **Orders**.



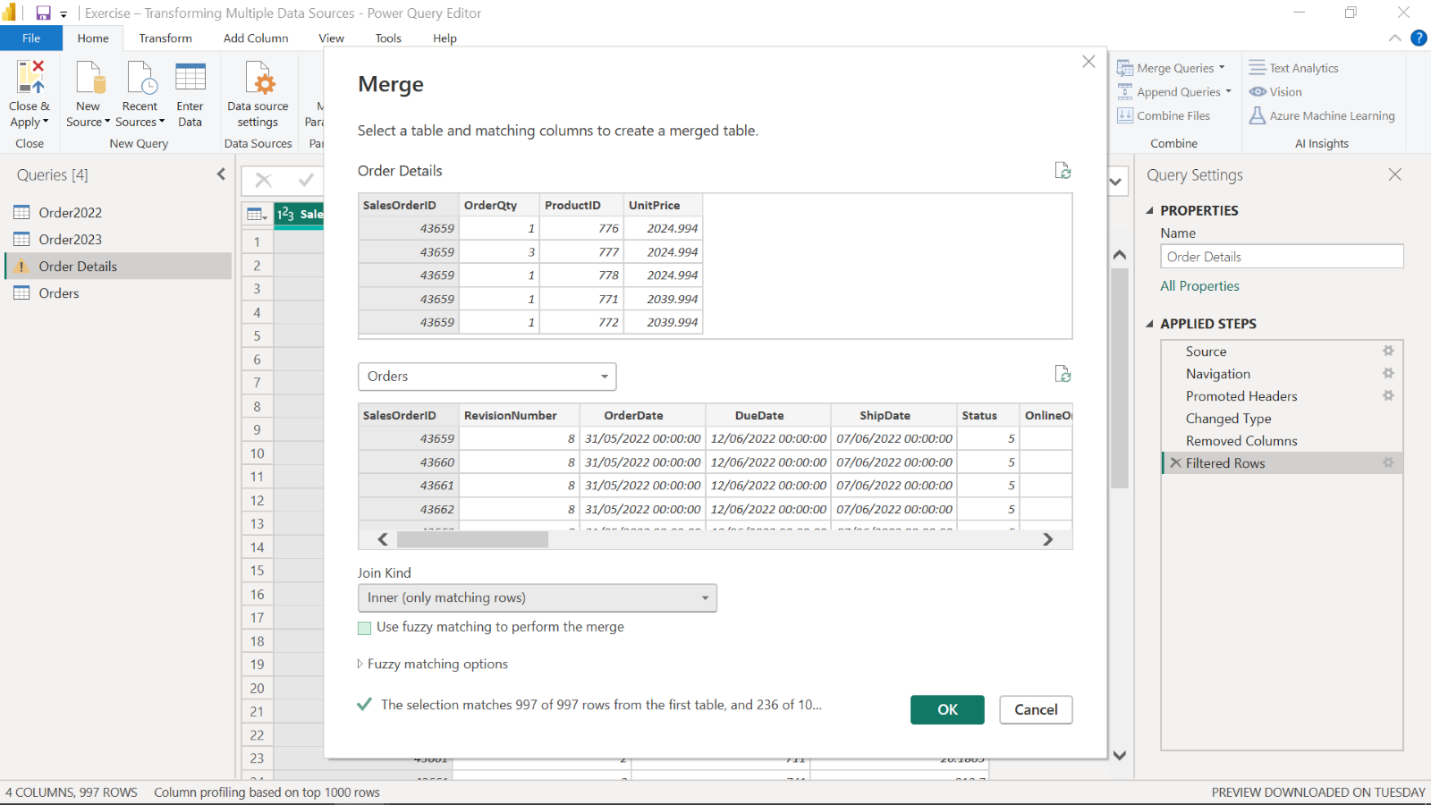
**Step 7: Merge Queries:**

In the final step, you are asked to merge the tables **OrderDetails** and **Orders** by using the common column between the tables and take the needed columns from each table.

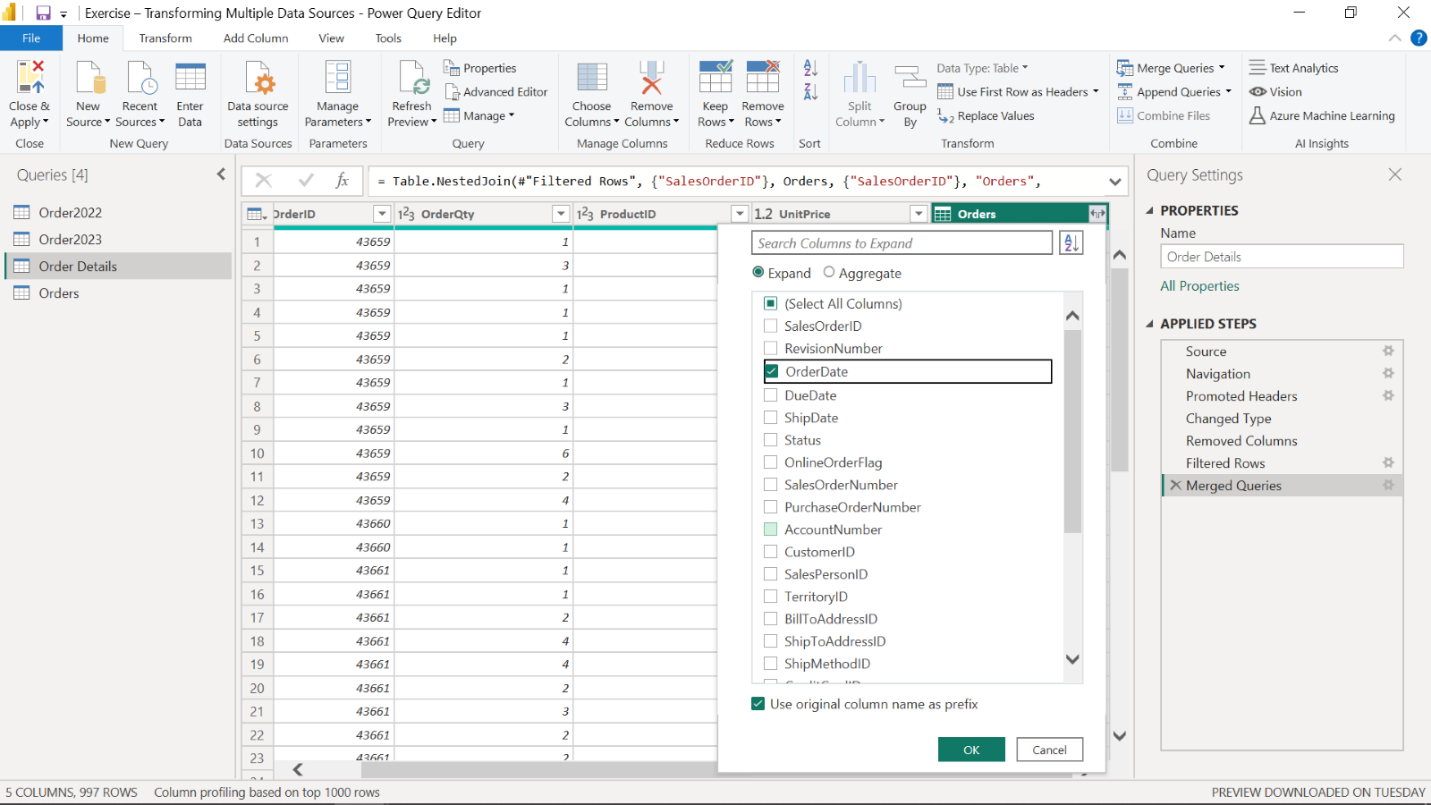
1. You selected the **OrderDetails** data in the Queries pane and from **Home** > **Combine** you chose **Merge Queries** in the **Merge Queries** drop-down menu.



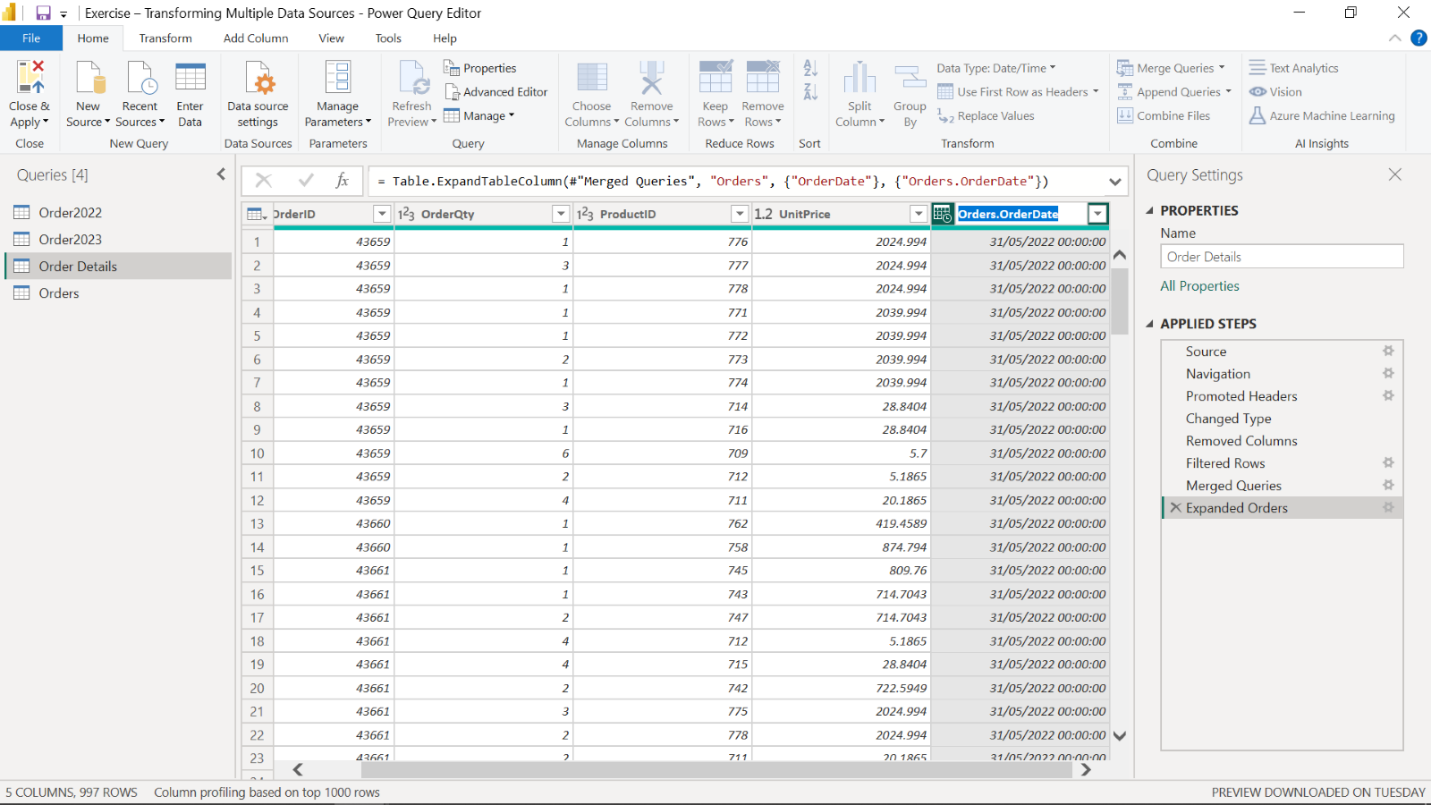
1. In the opened window, **OrderDetails** table was automatically shown in the upper part.
2. You selected the next table for merging, which was **Order** using **SalesOrderID** as the common column between the tables to establish connection.



1. For the **Join Kind** dropdown, you selected the join type of **Inner Join**, which selects the matching records from the left table and the right table.
2. Then you selected **Append** near the newly added **Orders** table column and selected only the **OrderDate** column from **the Orders** table in the opened window.

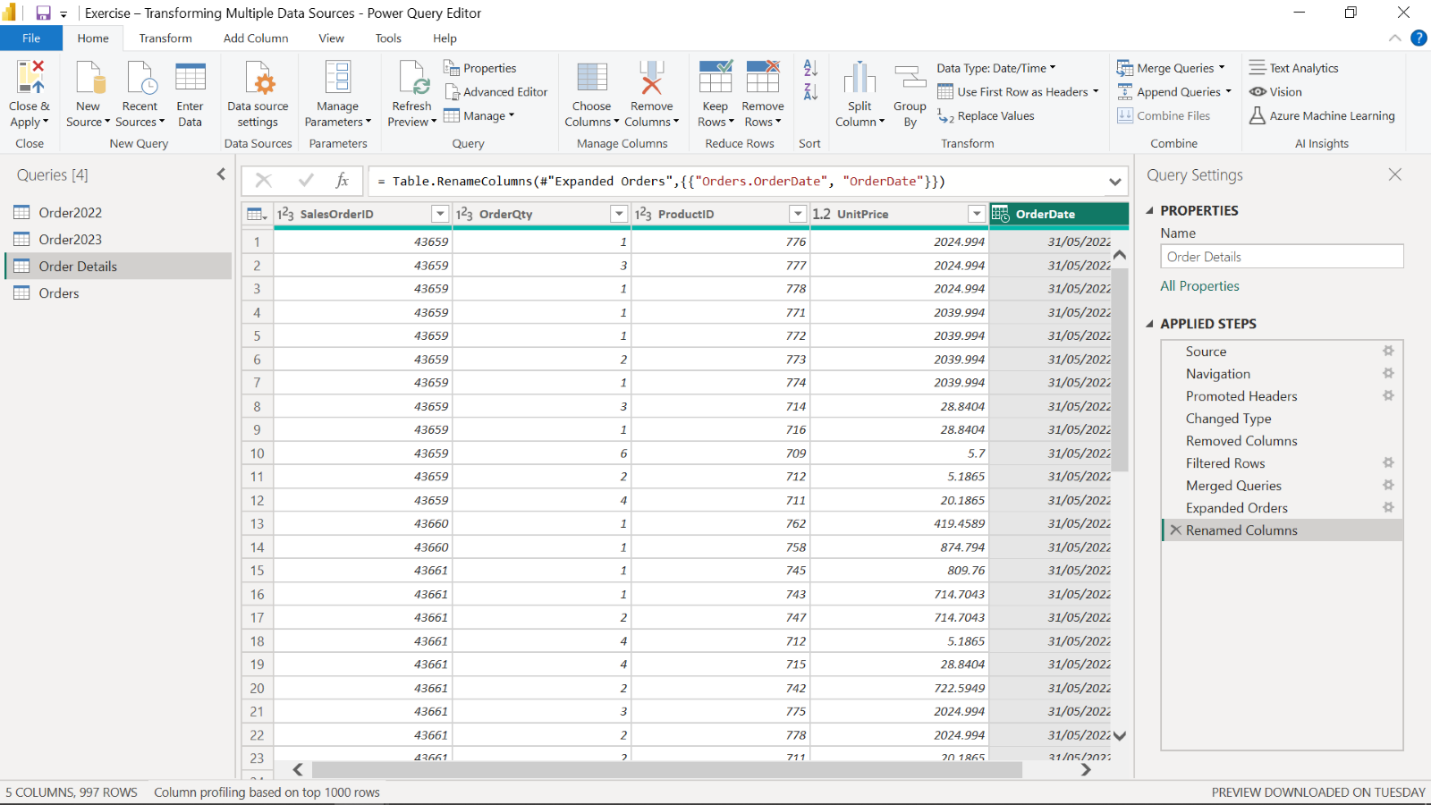


1. Using a double-click you then renamed the **Orders.OrderDate** column to its new name of **OrderDate**.



**Completed exercise**

Your final worksheet should look like this:



**Conclusion**

Your objective for this exercise was to complete an end-to-end scenario. In this context, you learned how to clean and transform multiple data sources, append, join and merge them and identify the potential anomalies using Power Query.