**Exercise: Preparing a dataset**

[SalesFile](https://d3c33hcgiwev3.cloudfront.net/jOajn-j3Qv61AgoNLrCDKg_3ab6df9b3f6a4ab393ce99aa917e7ba1_SalesFile.xlsx?Expires=1709769600&Signature=ku6M8S1Zy-QHF~YZRSdFER10dULbM54yTraUSz0QC1wGYcc8OM-1F77bkntAOzrzsaRZJ9WTOZmPKZdqJB-uJuos5~leXEiMMMf98pXf70QP~2sWeZ8cF8-3rwOL9Z7Dyva2cxP4OlD-Ui8szCQxkTwhBwo6i8FReXbzQ8~yRAY_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

[XLSX File](https://d3c33hcgiwev3.cloudfront.net/jOajn-j3Qv61AgoNLrCDKg_3ab6df9b3f6a4ab393ce99aa917e7ba1_SalesFile.xlsx?Expires=1709769600&Signature=ku6M8S1Zy-QHF~YZRSdFER10dULbM54yTraUSz0QC1wGYcc8OM-1F77bkntAOzrzsaRZJ9WTOZmPKZdqJB-uJuos5~leXEiMMMf98pXf70QP~2sWeZ8cF8-3rwOL9Z7Dyva2cxP4OlD-Ui8szCQxkTwhBwo6i8FReXbzQ8~yRAY_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

**Introduction**

By now, you should be familiar with common data errors and how to resolve them. You've explored several examples of common data errors and discovered how they can be resolved using Power Query. In this exercise, you'll put your knowledge into practice by importing, cleaning, and transforming data in Power BI Desktop to ensure accurate and reliable data analysis.

**Case study**

Adventure Works needs your help to analyze its sales data and create a Power BI report that visualizes this data in a meaningful way. Before the analysis can begin, the raw data must be cleaned and transformed to make sure it's accurate and consistent.

The company sends you the CSV file containing the raw dataset named *SalesFile.csv*. The dataset consists of important data related to recent sales, such as product categories, manufacturing prices, sales prices, units sold, and other similar details.

However, the file also contains errors like missing values, incorrect data types, and inconsistent formatting. Help Adventure Works to resolve these issues using Power Query editor so that they can produce an accurate and reliable report.

1. This exercise aims to assist you in understanding how to address common data issues such as missing values, incorrect data types, and inconsistent formatting.
2. By the end of this exercise, you’ll understand how to import, clean, and transform data in Power BI Desktop to ensure accurate and reliable analysis.

**Instructions**

Create a new Power BI **project** called **Exercise – Preparing a dataset.** Follow the prompts below to complete the exercise.

**Step 1: Load the workbook**

1. Download the Microsoft Excel workbook **SalesFile.xlsx.**
2. Import the **SalesFile.xlsx** Excel file as your dataset in Power BI.

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

1. Open thePower Query editorto begin editing your data.

**Step 3: Address missing values**

1. Locate and select the **Units Sold** column.
2. Identify all **null** values within the column and replace them with a value of **0**.
3. Repeat this task for the **Sale Price**, **Sales**, and **Profit** columns.

**Step 4: Clean the Manufacturing Price and Sale Price columns**

1. Locate and select the **Manufacturing Price** and **Sale Price** columns.
2. Change the data type for both columns to **Decimal Number**.
3. Repeat this task for the **Sales** and **Profit** columns.

**Step 5: Clean the Discount Band Column**

1. Select the **Discount Band** column.
2. Locate each instance of value **1** in the column. Replace each instance of this value with **None**.
3. Then change the data type of the column to **Text**.

**Step 6: Clean the Units Sold column**

1. Select the **Units Sold** column. Search for and locate all instances of the text value **six hundred**.
2. Replace each instance of this text value with the numerical value **600**.
3. Then change the column's data type to **Whole Number**.

**Step 7: Address inconsistencies in the Date column**

1. Select the **Date** column. Ensure that the column’s data type is **Date**.
2. The column also contains several null values. Replace all null values with the default date of **March 03rd 2023**.
3. Next, select the **Month Number** column. Change the column's data type to **Whole Number**.

**Step 8: Drop records with errors**

1. Select the **Manufacturing Price** column. The column contains errors in rows **6** and **38**. Use the **Remove Errors** feature to drop these records.
2. Repeat the same steps for the errors in the **Sales** and **Profit** columns.

**Step 9: Drop duplicate rows**

1. You need to identify several duplicate rows that the dataset contains.  **Tip:** Check the Products column
2. Once identified, remove these duplicate rows from the dataset.

**Step 10: Apply the data transformations**

1. Once you have completed all the above data cleaning steps, select the **Close & Apply** button.
2. By completing this action, you can apply your transformations and update the dataset.

**Conclusion**

By following these steps, you've successfully cleaned and transformed your dataset using Power BI. The dataset is now ready for further analysis and visualization in Power BI Desktop.

Remember, investing time and effort in cleaning your data will greatly enhance the quality of your analysis and ultimately lead to better decision-making and business outcomes.

Always strive for accuracy and consistency, and don't be afraid to dive into the details when working with data. With these skills, you'll be well-equipped to tackle any data challenges that come your way!

**Exemplar: Preparing a dataset**

**Introduction**

In the exercise *Preparing a data set,* you put into practice your understanding of how to prepare a data set by importing, cleaning, and transforming data in Power Query.

Your objective in this exercise was to prepare a worksheet for analysis by completing the following tasks:

* Address instances of missing values.
* Clean columns by changing data types as required.
* Replace values where required.
* Address inconsistencies in data.
* And drop records with errors.

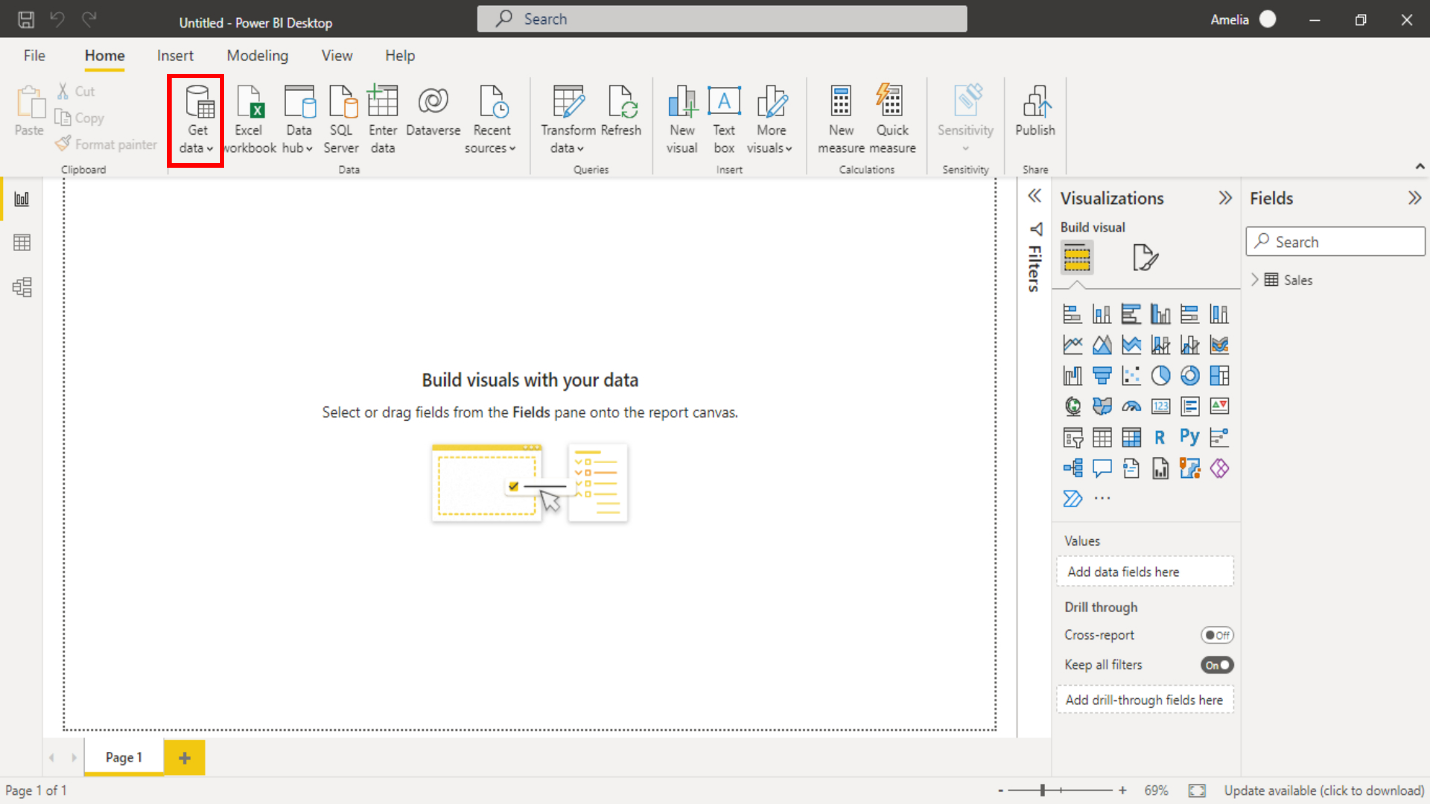
This reading provides a step-by-step guide for completing these tasks, accompanied by screenshots for easy comparison with your own copy. It also offers additional resources that provide more information on related topics.

**Step 1: Load the workbook**

*Import your dataset.*

Download the Microsoft Excel workbook *SalesFile.xlsx*.

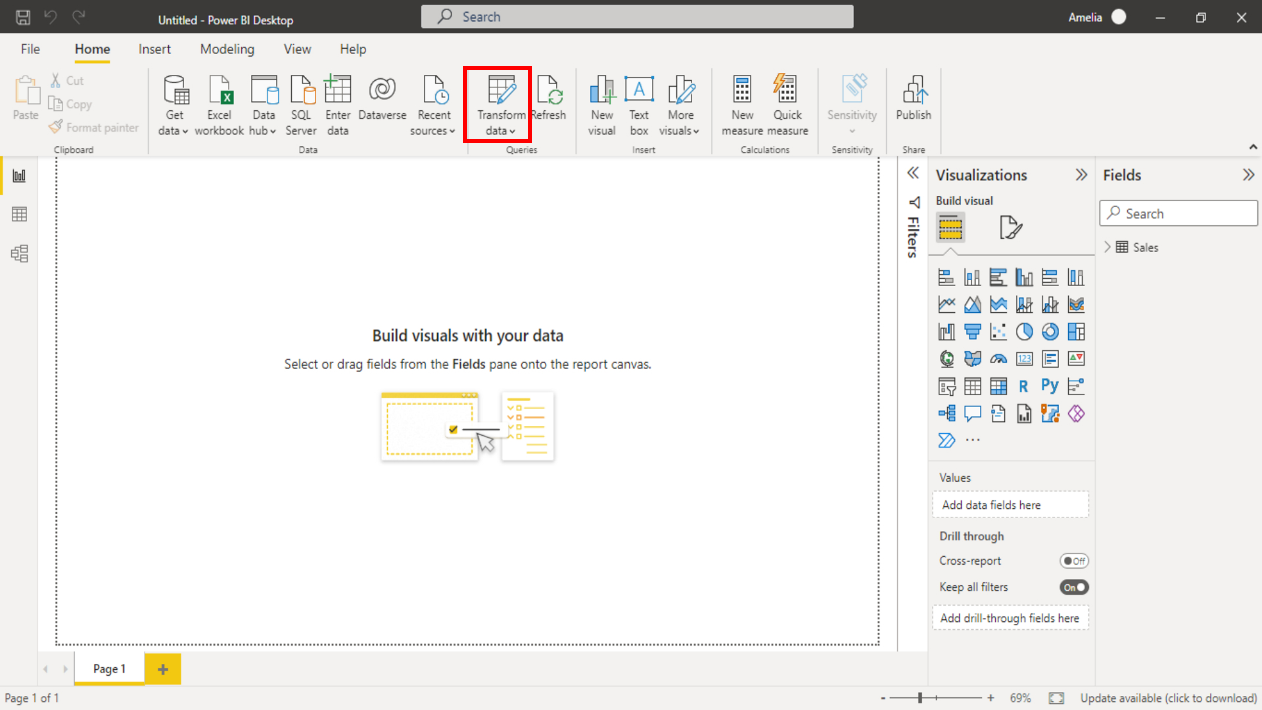
1. At the top of the **Power BI** window, navigate to the **Home** ribbon tab.
2. In the **Data** group, select the **Get Data** option. Select *SalesFile.xlsx* as the data source to import.



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

*Open the Power Query editor to begin editing your data.*

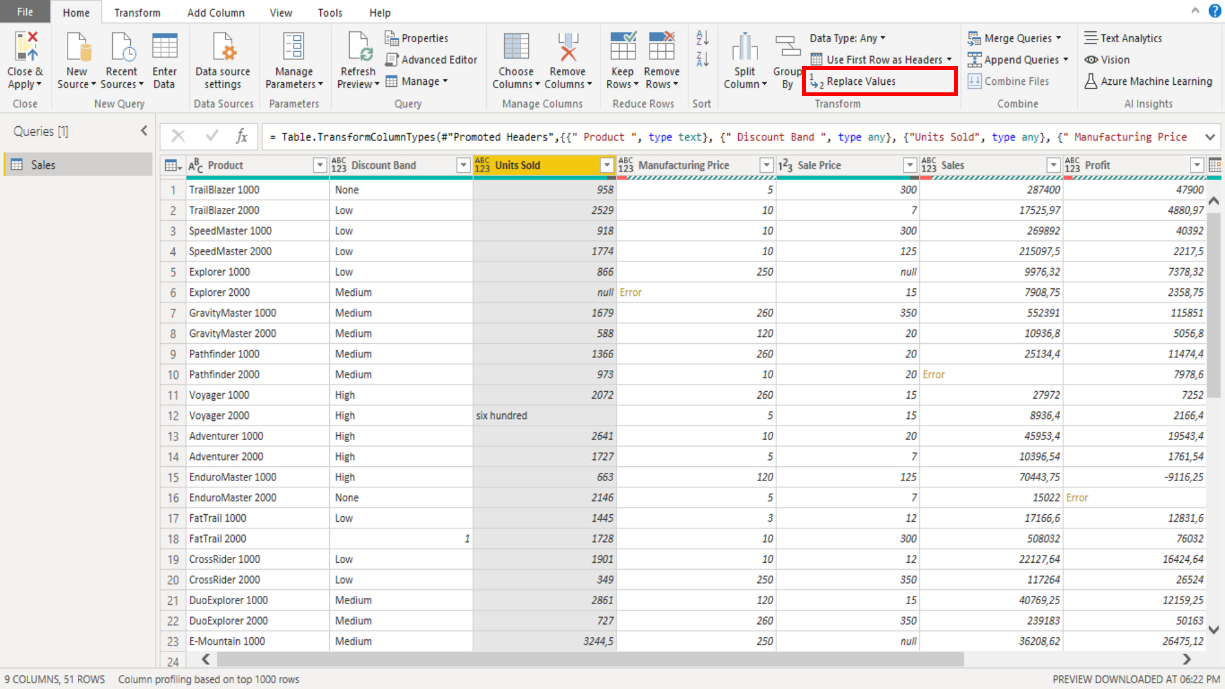
1. In the **Queries** group, select the **Transform Data** button. It’s located in the middle of the toolbar.
2. The **Power Query Editor** window opens. You can now begin cleaning the data.



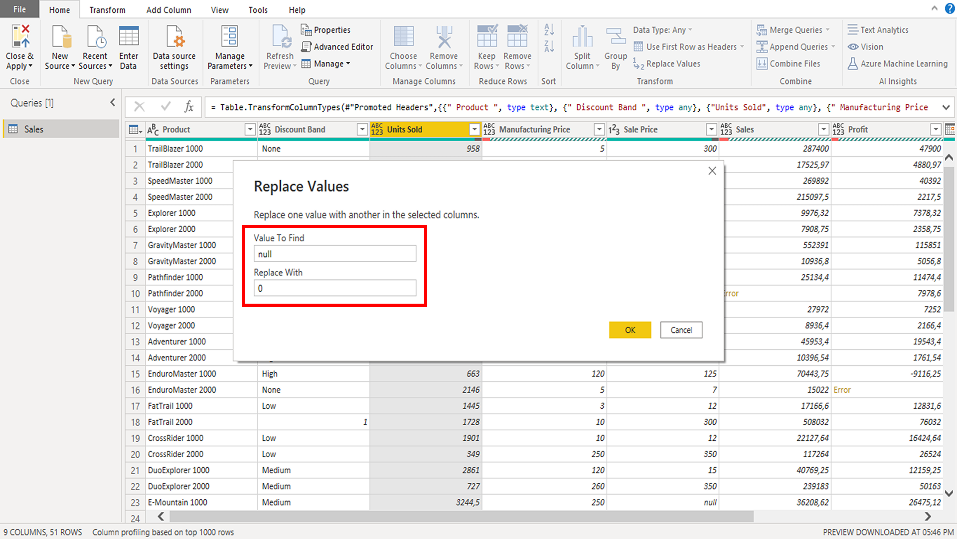
**Step 3: Address missing values**

*Locate and select the* ***Units Sold*** *column. Identify all* ***NULL*** *values within the column and replace them with a value of* ***0****. Repeat this task for the* ***Sale Price****,* ***Sales****, and* ***Profit*** *columns.*

1. In the **Power Query Editor**, locate the **Units Sold** column in the dataset preview.
2. Select the header of the **Units Sold** column to highlight it.
3. In the **Home** tab at the top of the window, select the **Replace Values** button.



1. In the **Replace Values** window that appears, type **null** in the **Value to Find** field (indicating missing values) and type **0** in the **Replace With** field.

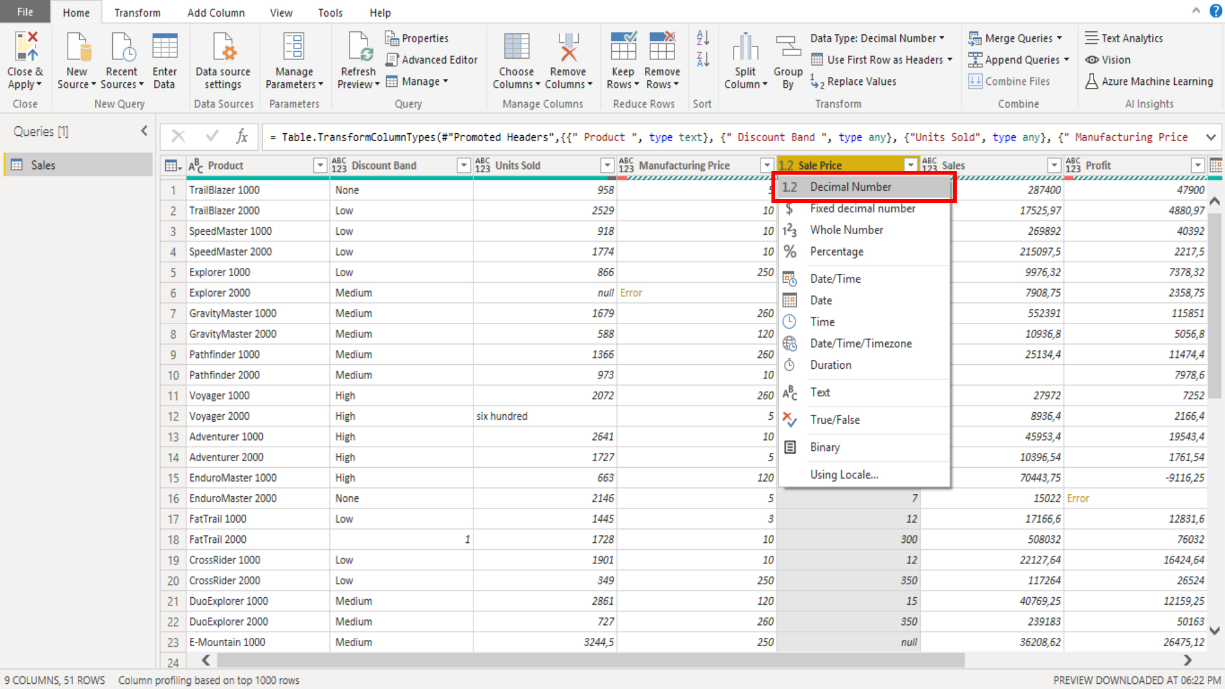


1. Select the **OK** button to replace all missing values in the **Units Sold** column with **0**.
2. Now, select the down arrow near the **Units Sold** column and confirm that no **NULL** values remain in the column. By performing this operation, you have replaced the non-numeric value **NULL** with the numeric value **0** in the **Units Sold** rows.
3. Repeat these steps for the **Sale Price**, **Sales**, and **Profit** columns.

**Step 4: Clean the Manufacturing Price and Sale Price columns**

*Locate and select the* ***Manufacturing Price*** *and* ***Sale Price*** *columns. Change the data type for both columns to* ***Decimal Number****.*

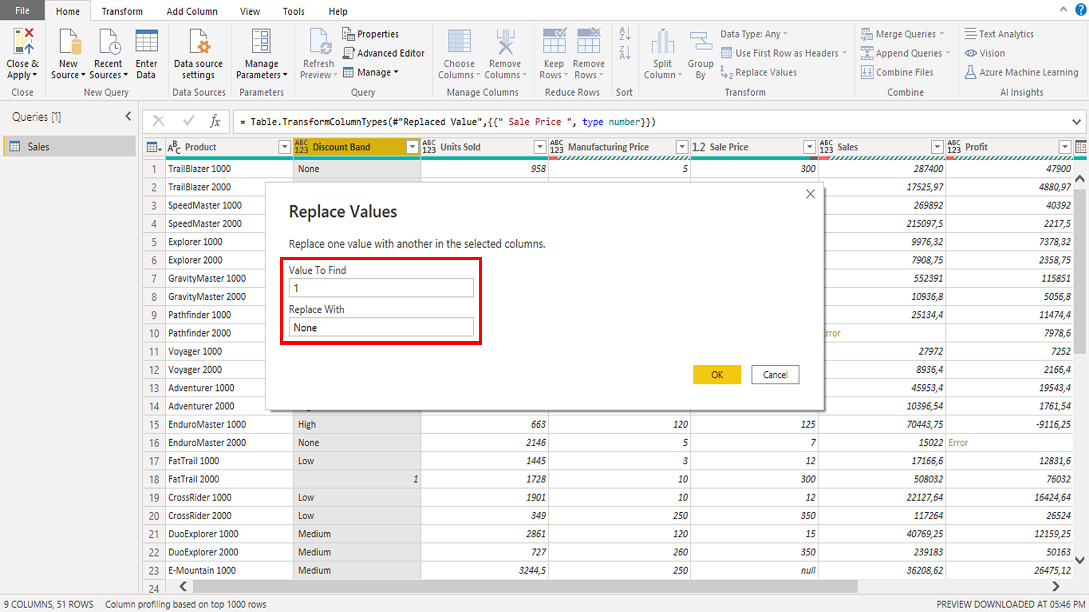
1. Highlight the **Manufacturing Price** column by selecting its header.
2. Change the data type of the **Manufacturing Price** column to **Decimal Number** by selecting the data type icon next to the column name (it may initially show as **ABC** or **123**) and choosing **Decimal Number** from the drop-down menu.
3. Take note of how the appearance of the **Manufacturing Price** column values has changed.
4. Repeat this task for the **Sale Price**, **Sales**, and **Profit** columns.



**Step 5: Clean the Discount Band column**

*Select the* ***Discount Band*** *column. Locate each instance of value* ***1*** *in the column. Replace each instance of this value with* ***None****. Then change the data type of the column to* ***Text****.*

1. Highlight the **Discount Band** column by selecting its header.
2. Select the **Replace Values** button in the **Home** ribbon tab.
3. In the **Replace Values** window, enter **1** in the **Value to Find** field and **None** in the **Replace With** field.

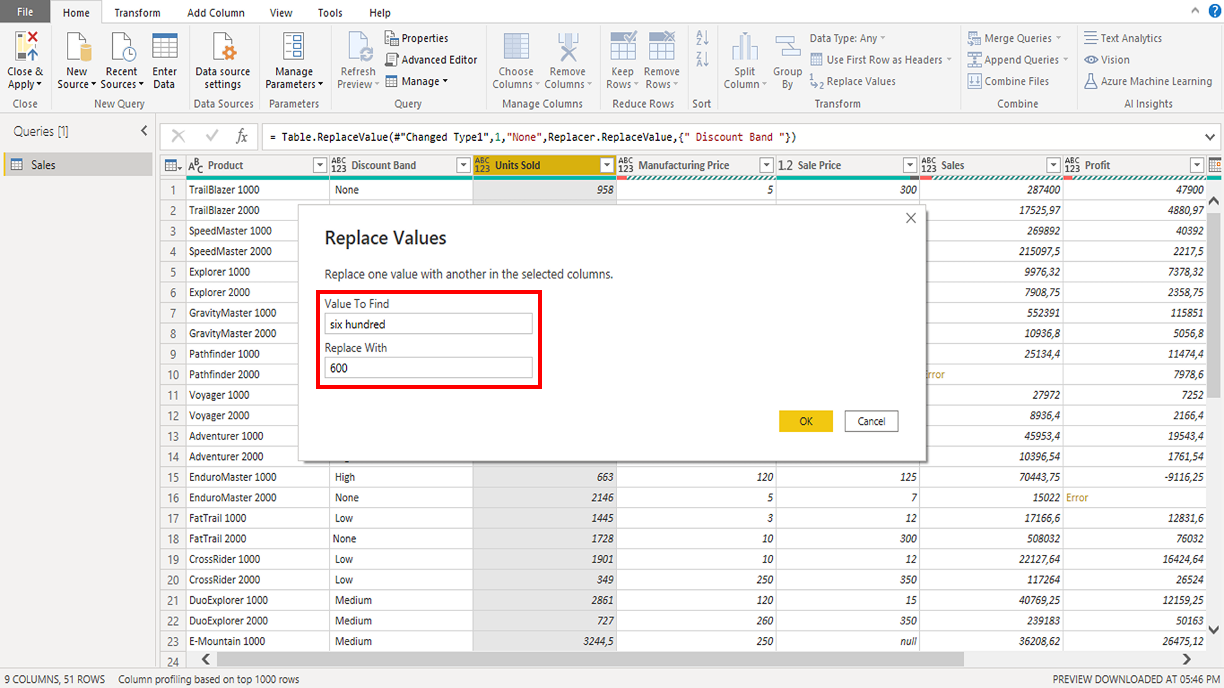


1. Select the **OK** button to replace the incorrect entry with **None**.
2. Select the down arrow near the **Discount Band** column and confirm that the **None** value is added to the value list.
3. Change the data type of the **Discount Band** column to **Text** by selecting the data type icon next to the column name (it may initially show as **ABC** or **123**) and choosing **Text** from the drop-down menu.
4. Take note of how the appearance of the **Discount Band** column values has changed.

**Step 6: Clean the Units Sold column**

*Select the* ***Units Sold*** *column. Search for and locate all instances of the text value* ***six hundred****. Replace each instance of this text value with the numerical value* ***600****. Then change the column’s data type to* ***Whole Number****.*

1. Highlight the **Units Sold** column by selecting its header.
2. On the **Home** tab, select the **Replace Values** button.
3. In the **Replace Values** window, enter **six hundred** (the text value to be replaced) in the **Value to Find** field and **600** (the new numeric value) in the **Replace With** field.

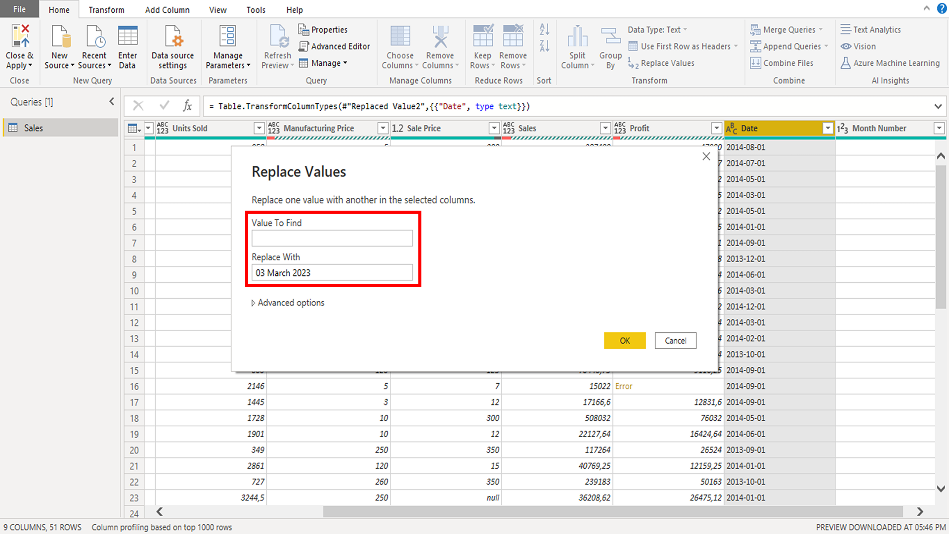


1. Select the **OK** button to replace the text entry with the numerical value. By performing this operation, you have replaced another non-numeric value of **six hundred** with a numeric value of **600** in the **Units Sold** rows. Now you are ready to change the column's data type.
2. Change the data type of the **Units Sold** column to **Whole Number** by selecting the data type icon next to the column name (it may initially show as **ABC** or **123**) and choosing **Whole Number** from the drop-down menu.
3. Take note of how the appearance of the **Units Sold** column values has changed.

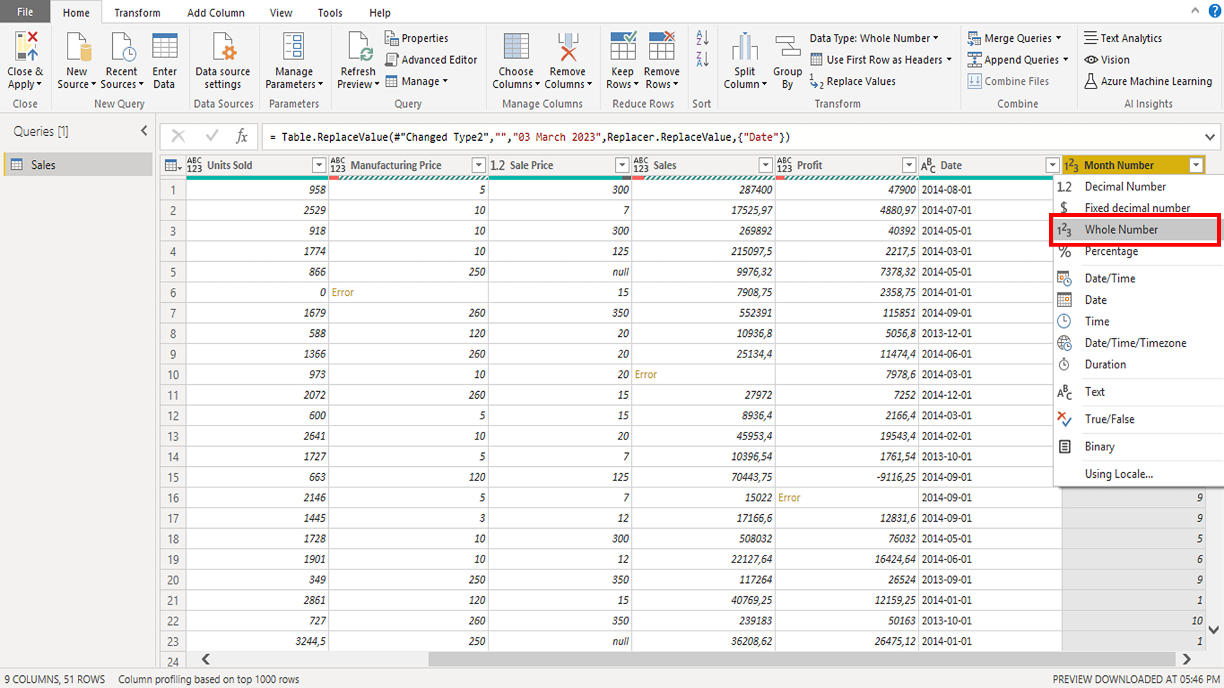
**Step 7: Address inconsistencies in the Date column**

*Select the* ***Date*** *column. Change the column’s data type to* ***Date****. The column also contains several empty values. Replace all empty values with the default date of* ***03 March 2023****. Next, select the* ***Month Number*** *column. Change the column’s data type to* ***Whole Number****.*

1. Highlight the **Date** column by selecting its header.
2. Change the data type of the **Date** column to the **Date** data type by selecting the data type icon next to the column name (it may initially show as **ABC** or **123**) and choosing the **Date** option from the drop-down menu.
3. Replace any empty values with a default date (of **03 March 2023**) using the **Replace Values** tool.



1. Next, select the down arrow near the **Month Number** column. Confirm that there are no empty values and that the **03 March 2023** value is added to the value list.
2. Highlight the **Month Number** column by selecting its header.
3. Change the data type of the **Month Number** column to **Whole Number** by selecting the data type icon next to the column name (it may initially show as **ABC** or **123**) and choosing **Whole Number** from the drop-down menu.

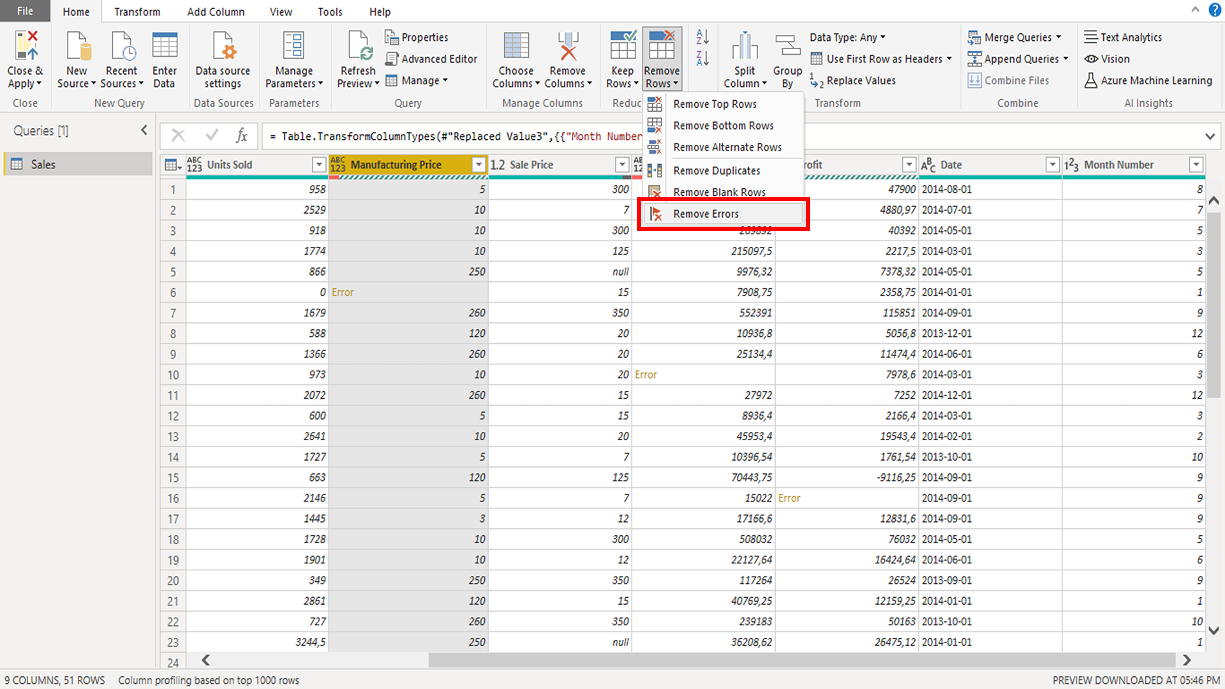


1. Take note of how the appearance of the **Month Number** column values has changed.

**Step 8: Drop records with errors**

*Select the* ***Manufacturing Price*** *column. The column contains errors in rows* ***6*** *and* ***38****. Use the* ***Remove Errors*** *feature to drop these records. Repeat the same steps for the errors in the* ***Sales*** *and* ***Profit*** *columns.*

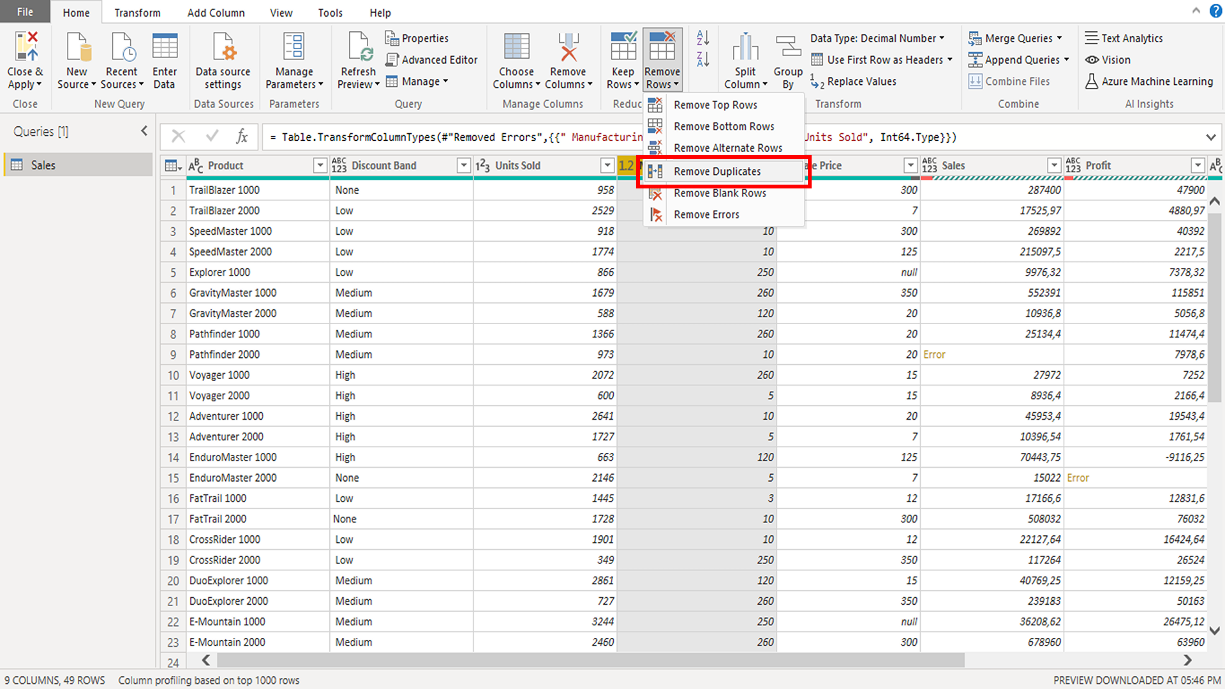
1. Examine each column individually to identify any columns that may have errors. There is an error in the **Manufacturing Price** column. Select the column header to select it.
2. With the column selected, navigate to the **Home** tab on the Ribbon.
3. Locate the **Remove Rows** button. Select the button to reveal a drop-down menu.
4. In the drop-down menu, choose **Remove Errors**. Power Query scans the selected column and removes all rows it identifies within the column that contain errors.
5. Confirm that no errors remain on any rows in the **Manufacturing Price** column.
6. Repeat these steps for the **Sales** and **Profit** columns.



**Step 9: Drop duplicate rows**

*The dataset contains several duplicate rows. Remove these duplicate rows from the dataset.*

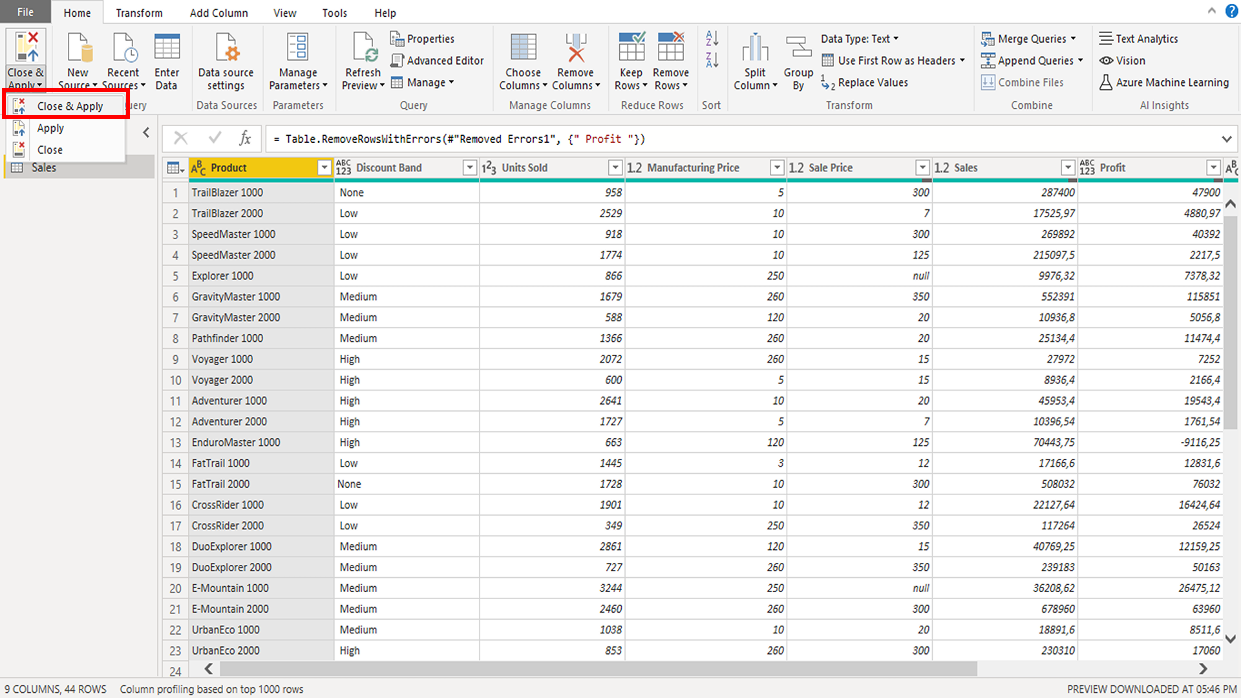
1. In the Power Query Editor, locate the **Queries** pane on the left side of the window. Select the table name.
2. In the **Home** tab, locate the **Remove Rows** button in the toolbar. Select the button to reveal a drop-down menu.
3. In the drop-down menu, choose **Remove Duplicates**. Power Query scans your dataset and automatically removes any duplicate rows it identifies.



**Step 10: Apply the data transformations**

*Once you have completed all the above data cleaning steps, select the* ***Close & Apply*** *button to apply your transformations and update the dataset.*

1. Once you have completed all the data cleaning steps outlined above, in the top-left corner of the **Power Query Editor** window, select the **Close & Apply** button. This action applies all the transformations you have made and closes the Power Query Editor.



1. Power BI Desktop updates the dataset in the **Data** view with the cleaned, transformed data.



**Conclusion**

Your objective for this exercise was to apply techniques for importing, cleaning, and transforming data. In this context, you learned how to use Power Query Editor to change data types, replace missing or inconsistent values, remove duplicate rows, and eliminate rows with errors.