**Exemplar: Evaluating data for transformation**

**Introduction**

In the exercise *Evaluating data for transformation*, you were entrusted with the role of a data analyst for Adventure Works. You were required to evaluate and transform blocks of data from different datasets using Microsoft Excel.

More specifically, you were asked to:

* Evaluate and rectify inconsistencies within the *Sales* dataset, such as missing **TransactionIDs** and **SalesAmounts**, and textual entries in numeric columns.
* Analyze the *Inventory* dataset, handle missing **ProductNames** and **Categories**, and standardize RestockingFrequency.
* Review the *Customer Feedback* dataset, correct irregularities in **FeedbackScores**, fix erroneous dates, and handle missing feedback entries.

This reading provides you with a guide that you can use to compare to your own work.

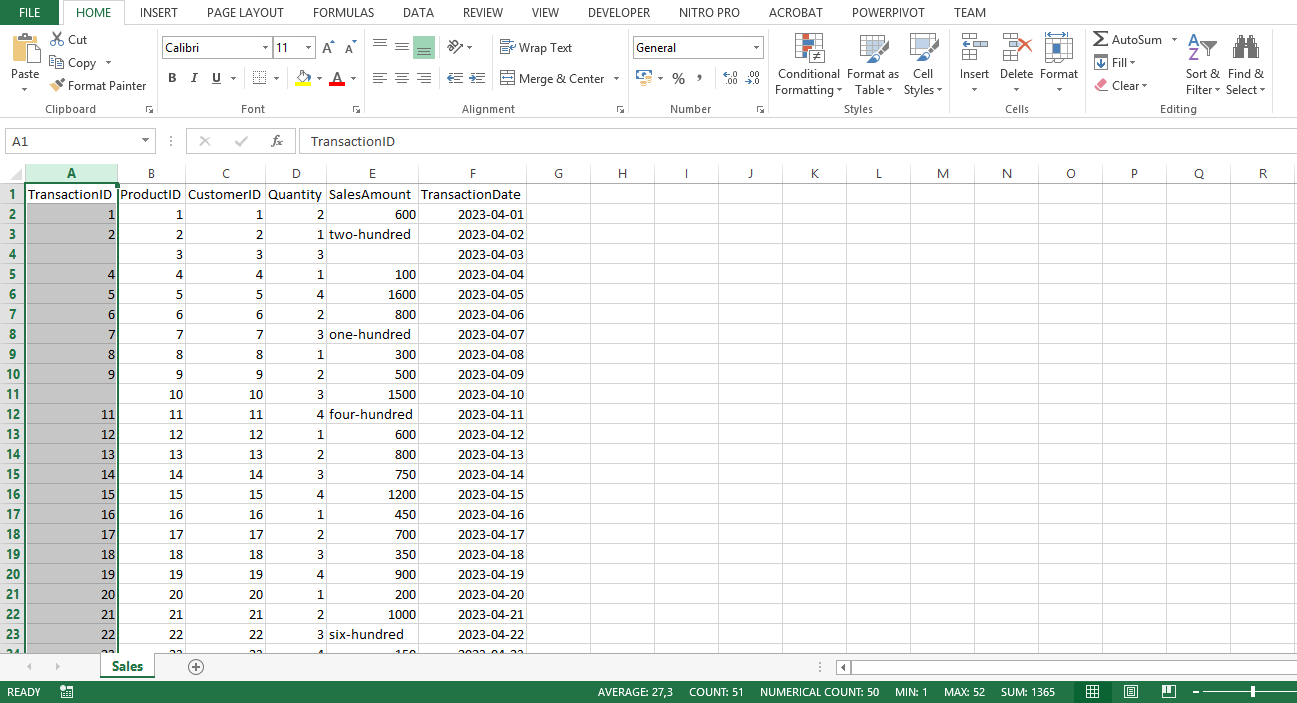
**Exemplar: Evaluating data for transformation**

1. You downloaded the *Sales Data***,***Inventory Data****,*** and *Customer Feedback Data* CSV filesto your computer.
2. You right-clicked the *Sales Data* CSV file and selected **Open with > Excel**.
3. You repeated this process for the other two files. This opened the files in Excel.

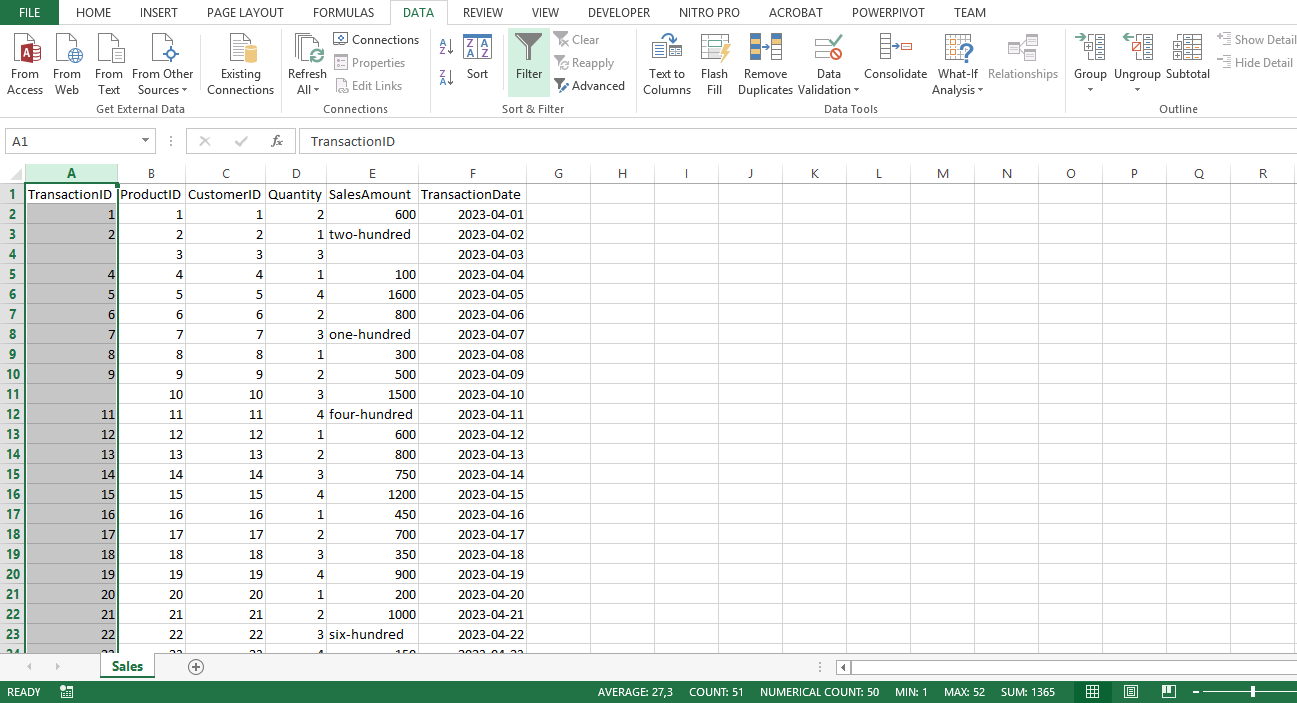
**Sales Dataset: What is the total sales value per product?**

**Step 1: Addressing Missing TransactionIDs**

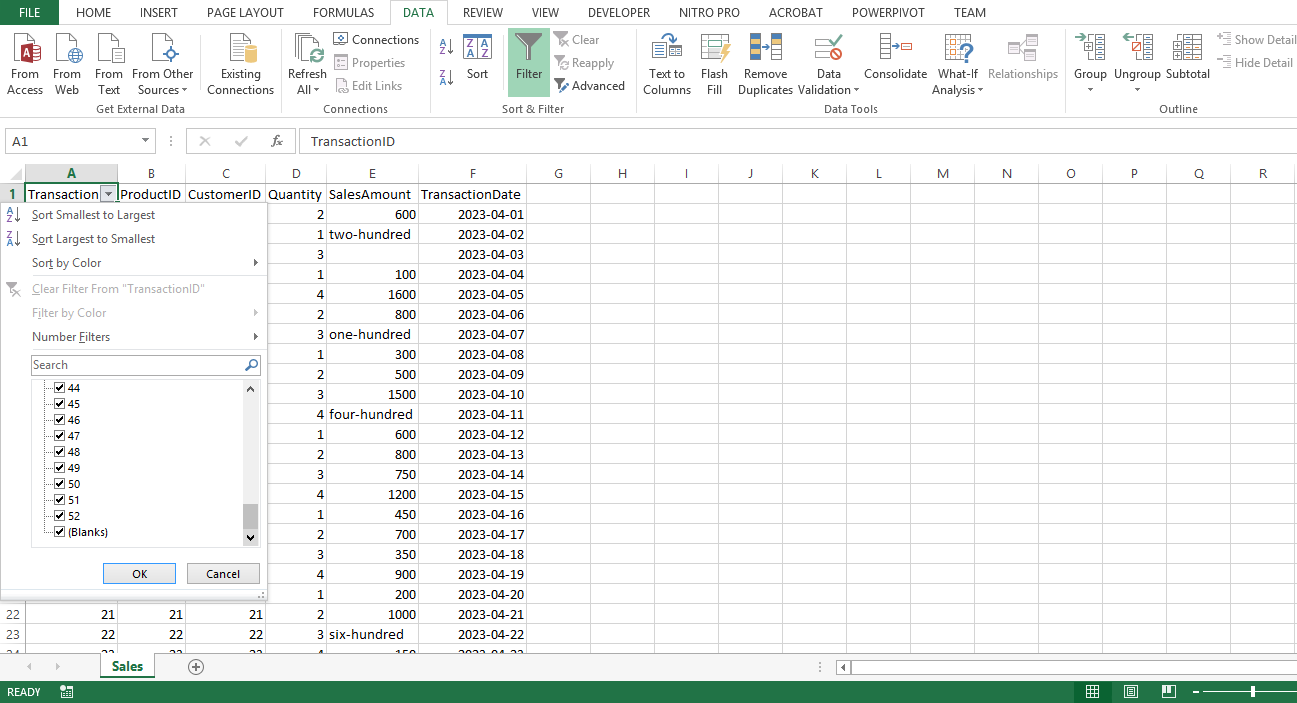
1. The initial step in transforming sales data was to deal with missing **TransactionIDs**. These are key identifiers, and their absence can skew the overall analysis. To handle this, you first selected the **TransactionID** column header to select the column.



1. In Excel on the **Data** tab, you selected **Filter**. This enabled filtering for your dataset.

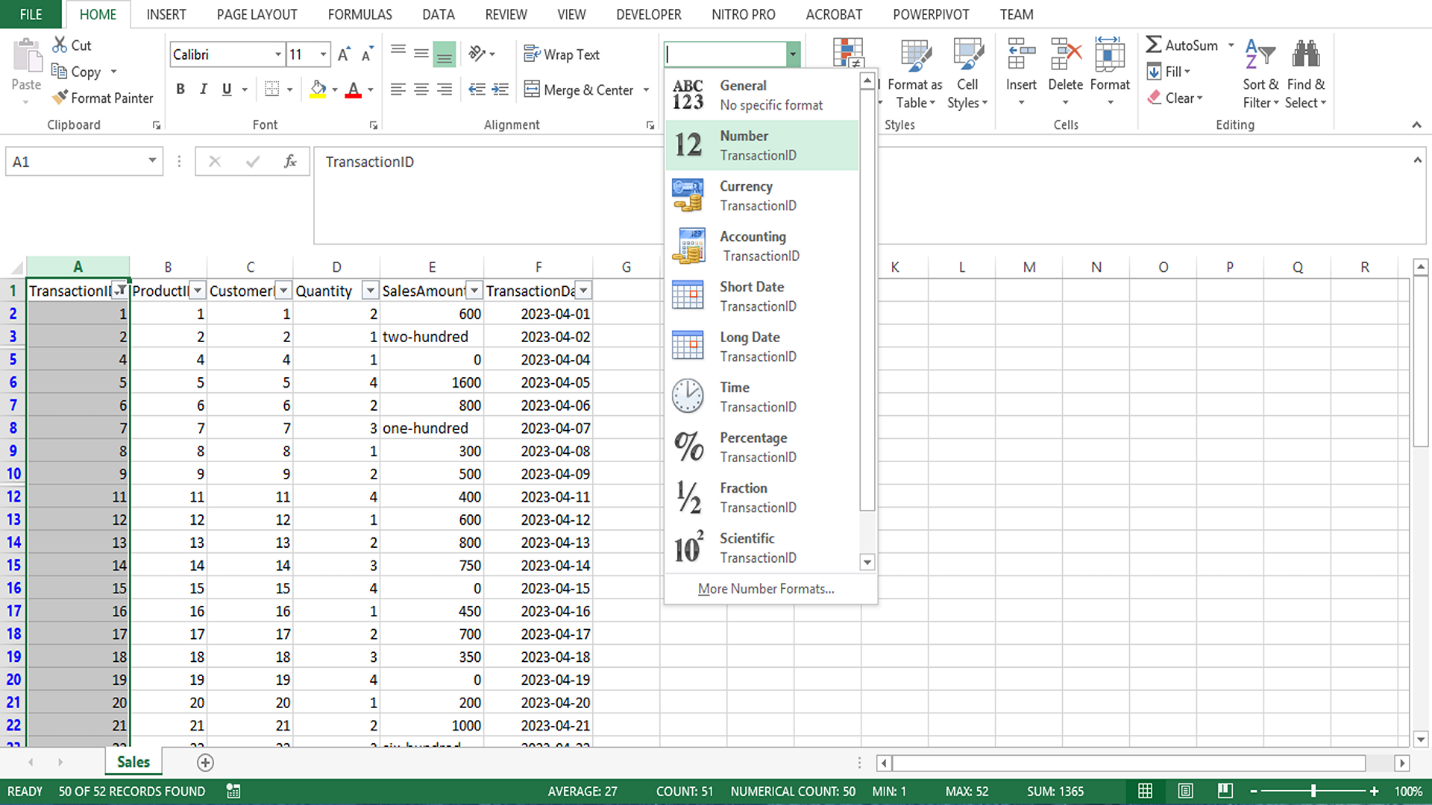


1. Selecting the drop-down arrow on the **TransactionID** column header displayed a menu with several options. You unchecked the box next to **Blanks** to exclude these entries.



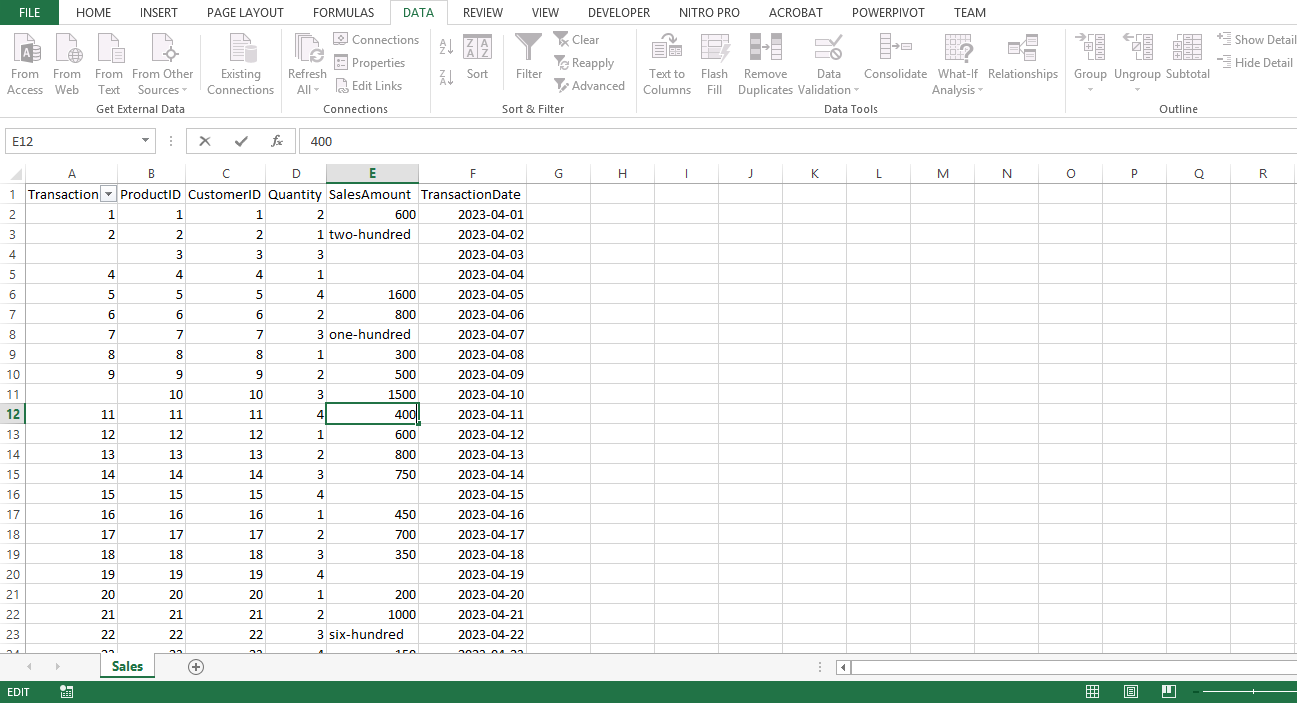
4. Selecting **OK** applied this filter.

5. Finally, you ensured that all the entries in the column are treated as numerical by setting the data type to **Number**.

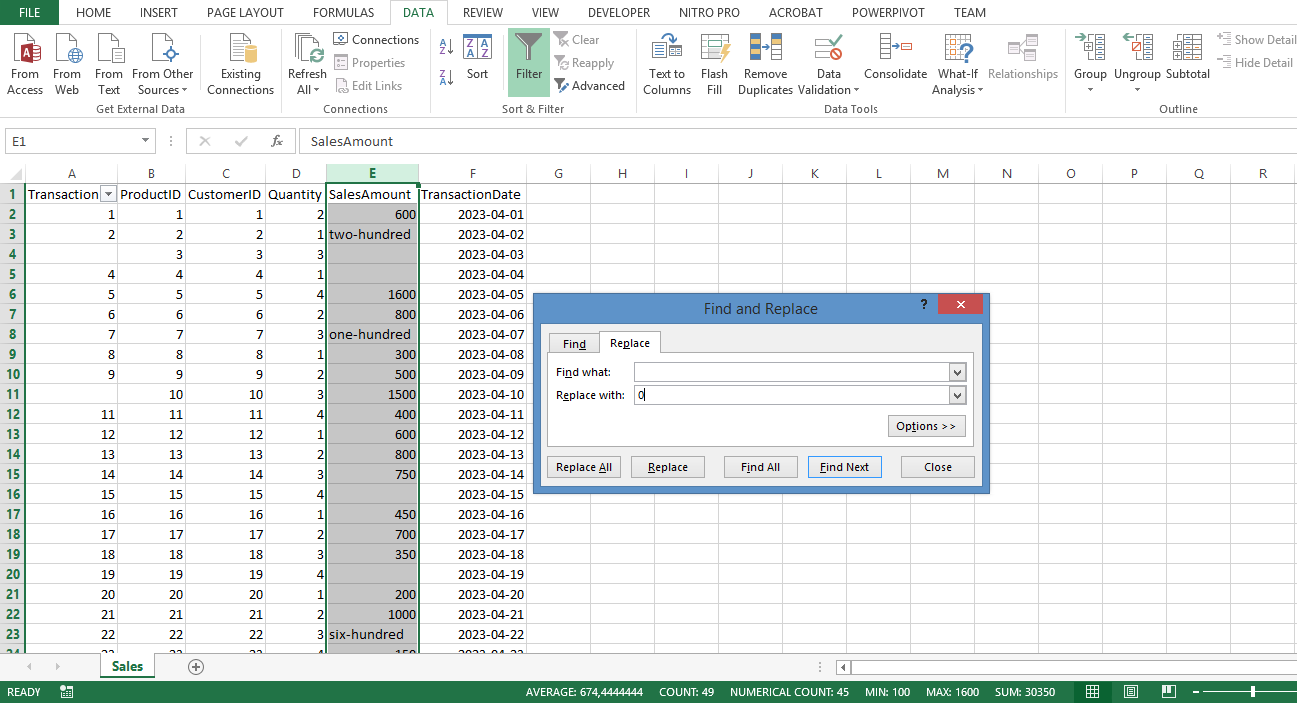


**Step 2: Dealing with SalesAmount anomalies**

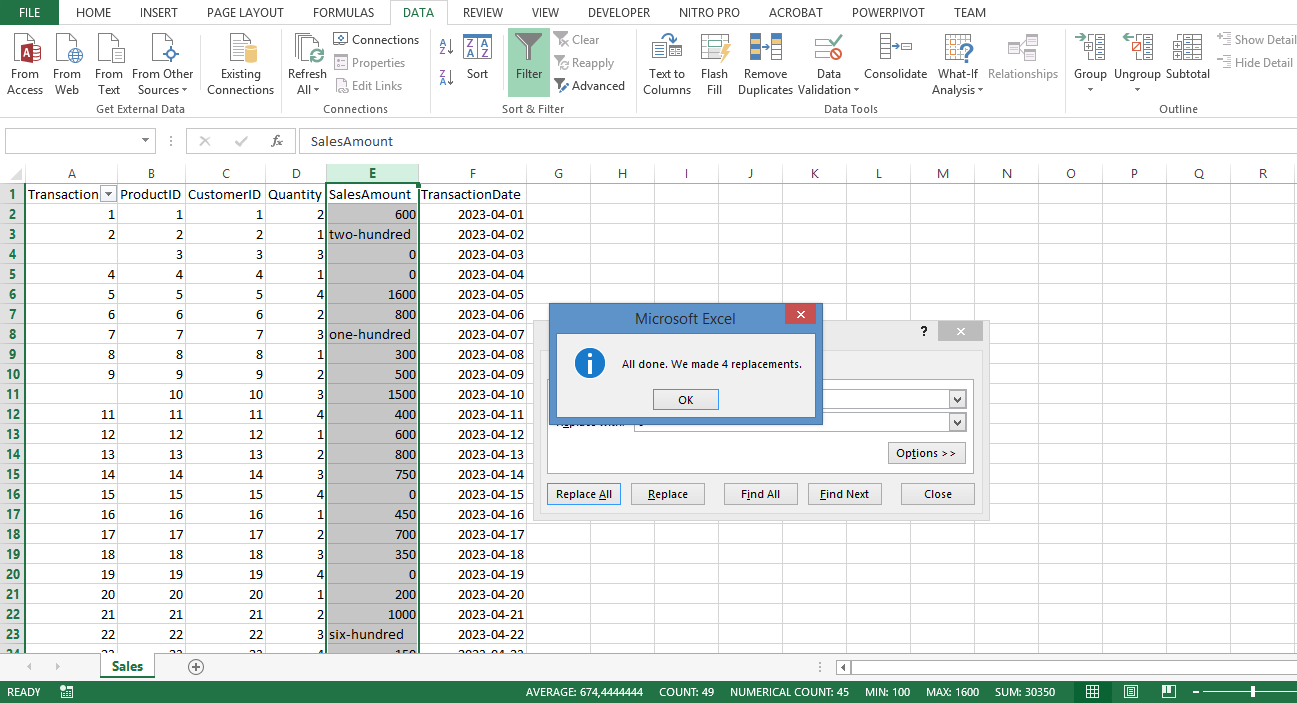
1. You encounter anomalies in the **SalesAmount** column, such as text values in place of numbers or missing values. You selected the **SalesAmount** column.
2. You manually checked the column to identify and replace non-numeric values such as four-hundred. To do this, you double-clicked the cell with the non-numeric value, deleted the existing value, and typed the appropriate numeric value.



1. Next, you replaced any missing **SalesAmounts** with **0** to prevent the exclusion of these entries from the analysis. To replace null or blank cells with **0**, you pressed **CTRL + H** to bring up the **Find and Replace** dialog box.



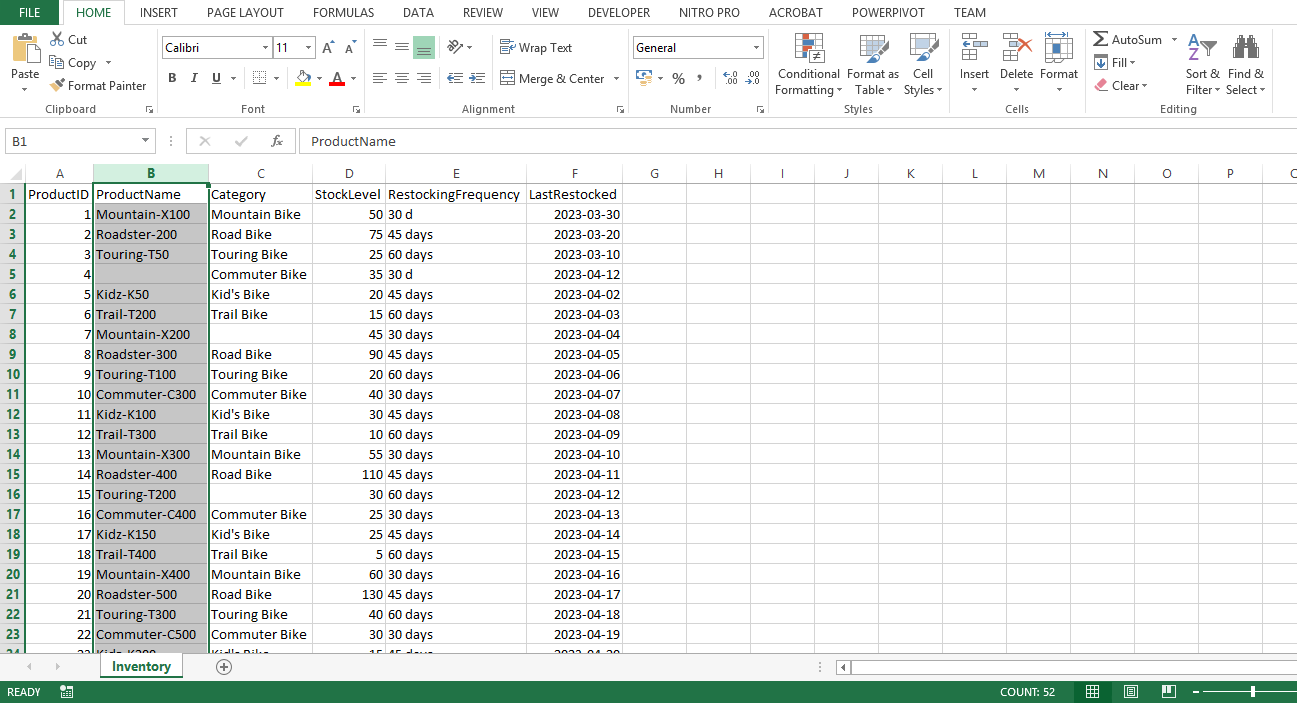
4. You left the **Find what** field, blank (this was to target all blank cells), and in the **Replace with** field, you typed **0**. You then selected **Replace All**. The Sales data was then clean and consistent, ready for further analysis in Power BI.



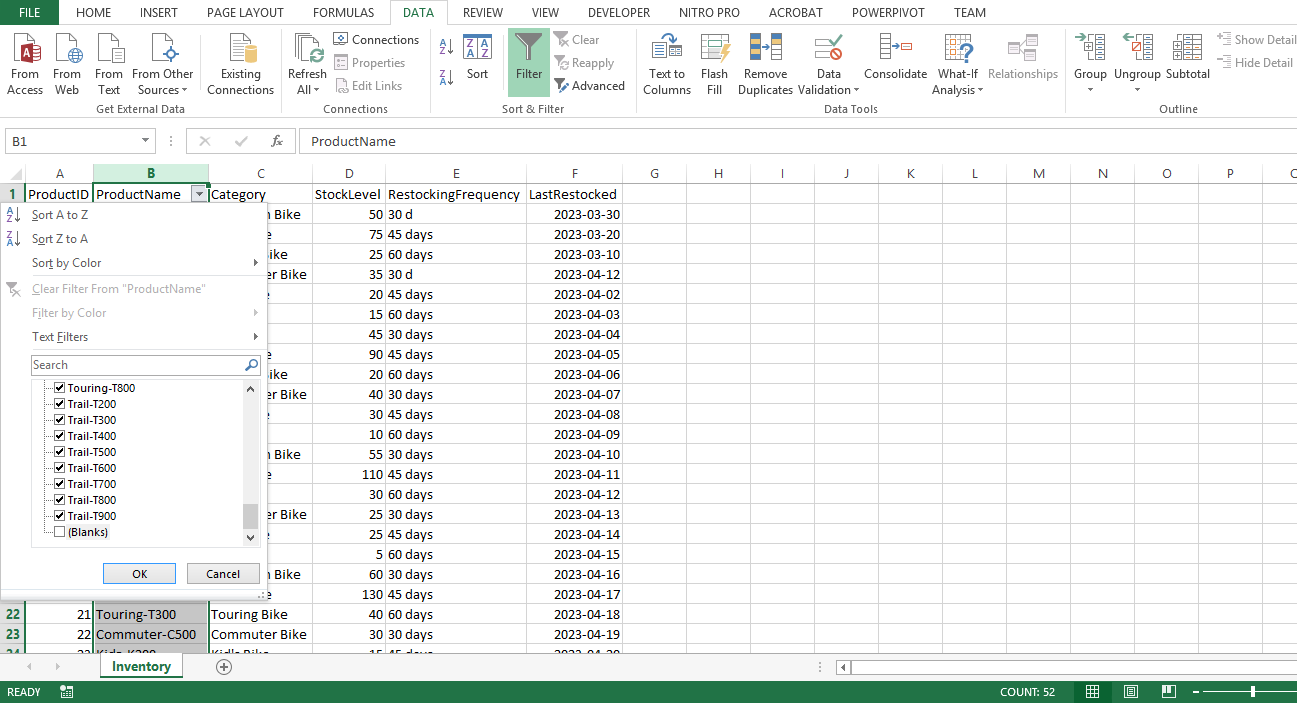
**Inventory Dataset: What is the stock level for each bike category?**

**Step 1: Addressing Missing ProductName**

1. You began by handling missing **ProductNames**. Similar to handling missing **TransactionIDs** in *Sales data*, you needed to filter out any rows with missing **ProductNames**, ensuring your analysis only includes complete records. You selected the **ProductName** column.

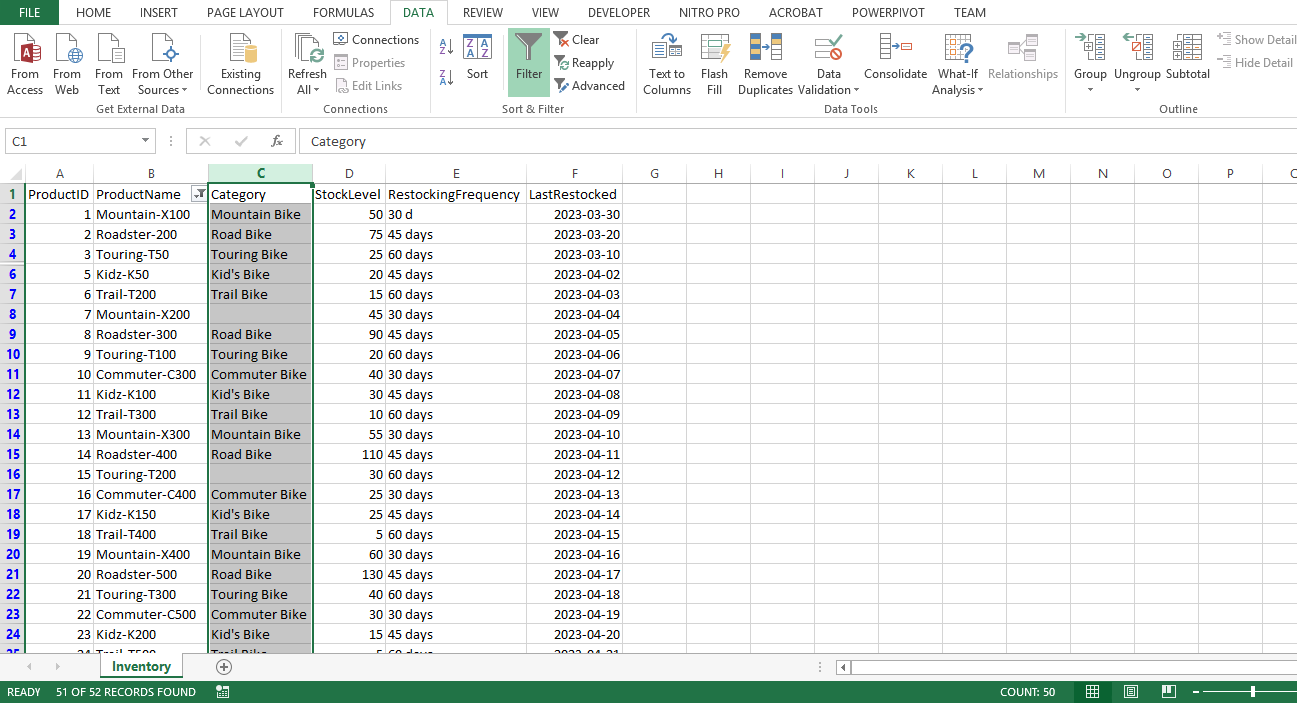


1. In the **Data** tab you selected **Filter**.
2. You selected the drop-down arrow on the **ProductName** column and deselected **Blanks**. Then selected **OK**.

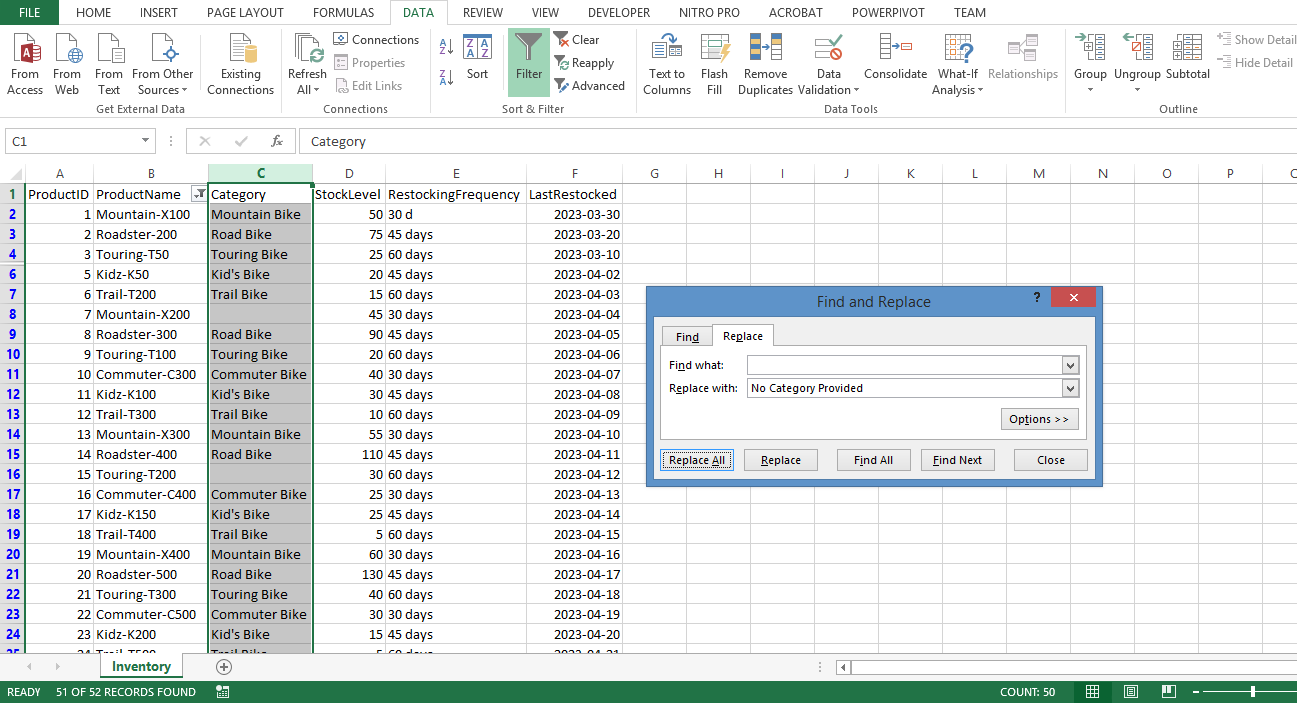


**Step 2: Handling missing Category**

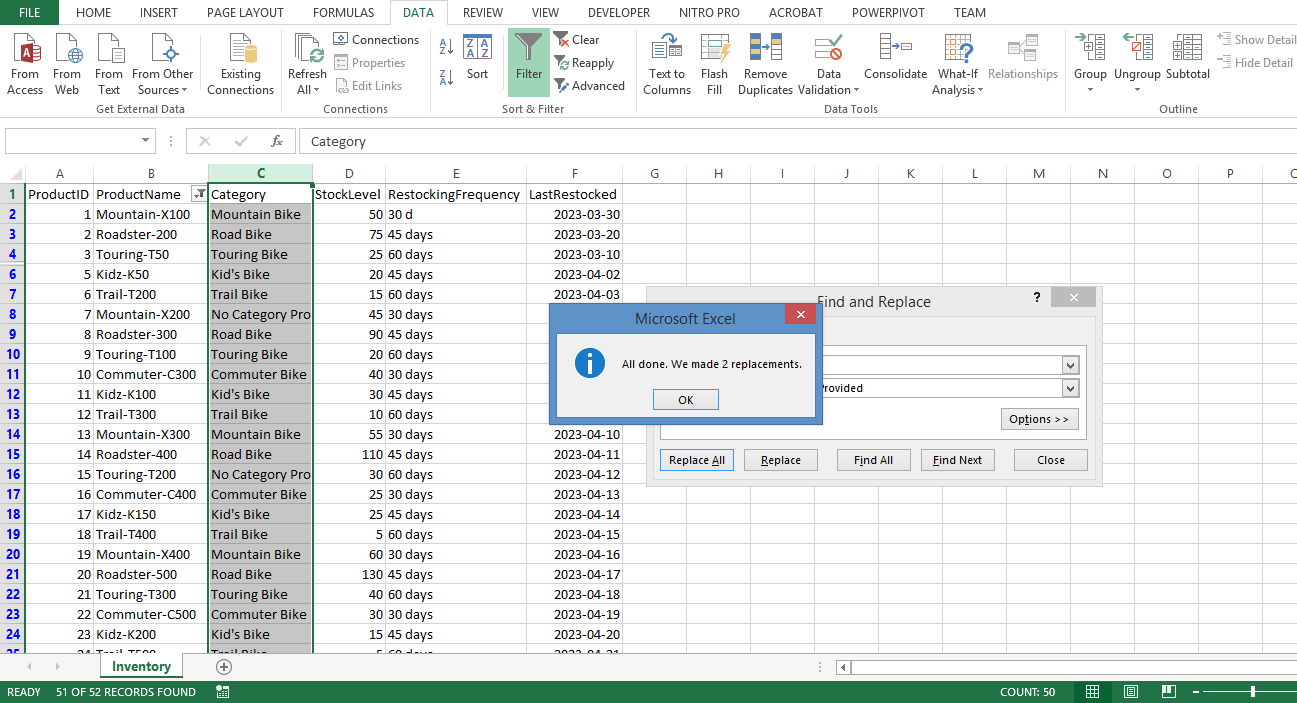
1. Next, you started evaluating the missing *Category* data, first selecting the **Category** column.
2. You manually checked the column and evaluated the missing entries.



1. To replace null or blank cells with **No Category Provided**, you pressed **CTRL + H** to bring up the **Find and Replace** dialog.

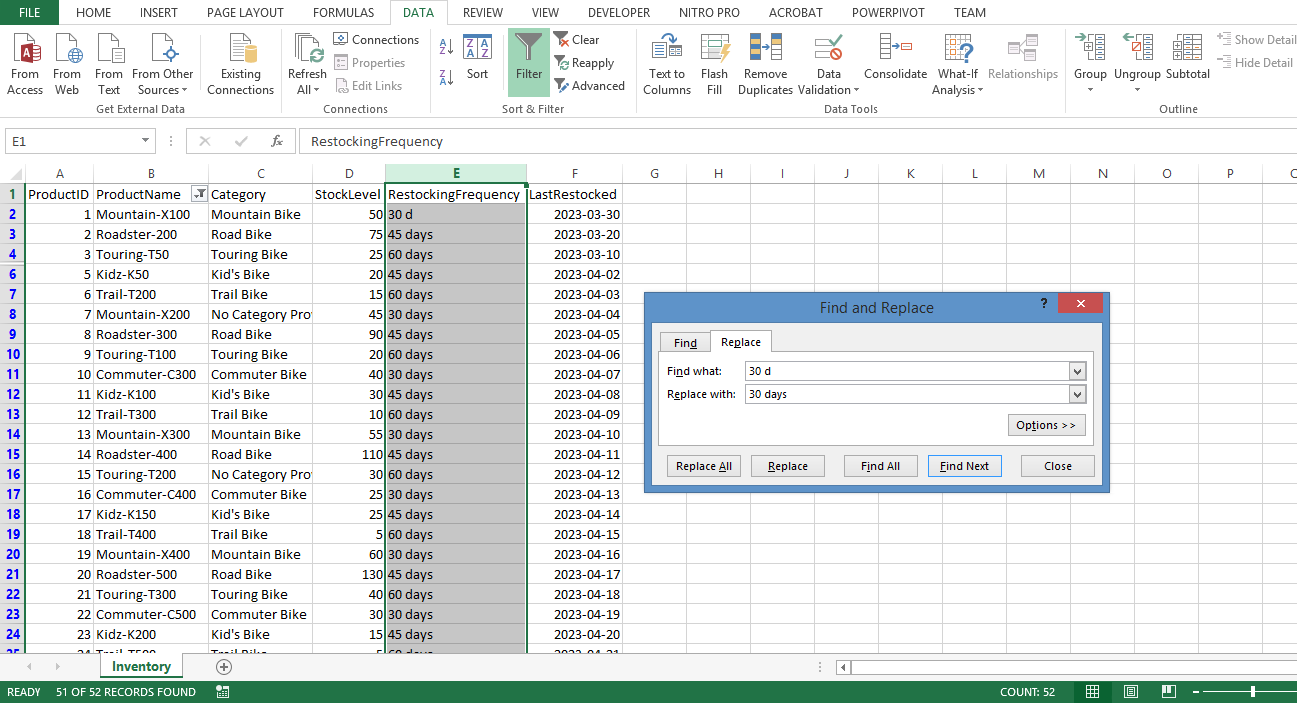


4. In the **Find what** field, you left it blank to target all blank cells, and in the **Replace with** field you entered **No Category Provided**. Then you selected **Replace All**.

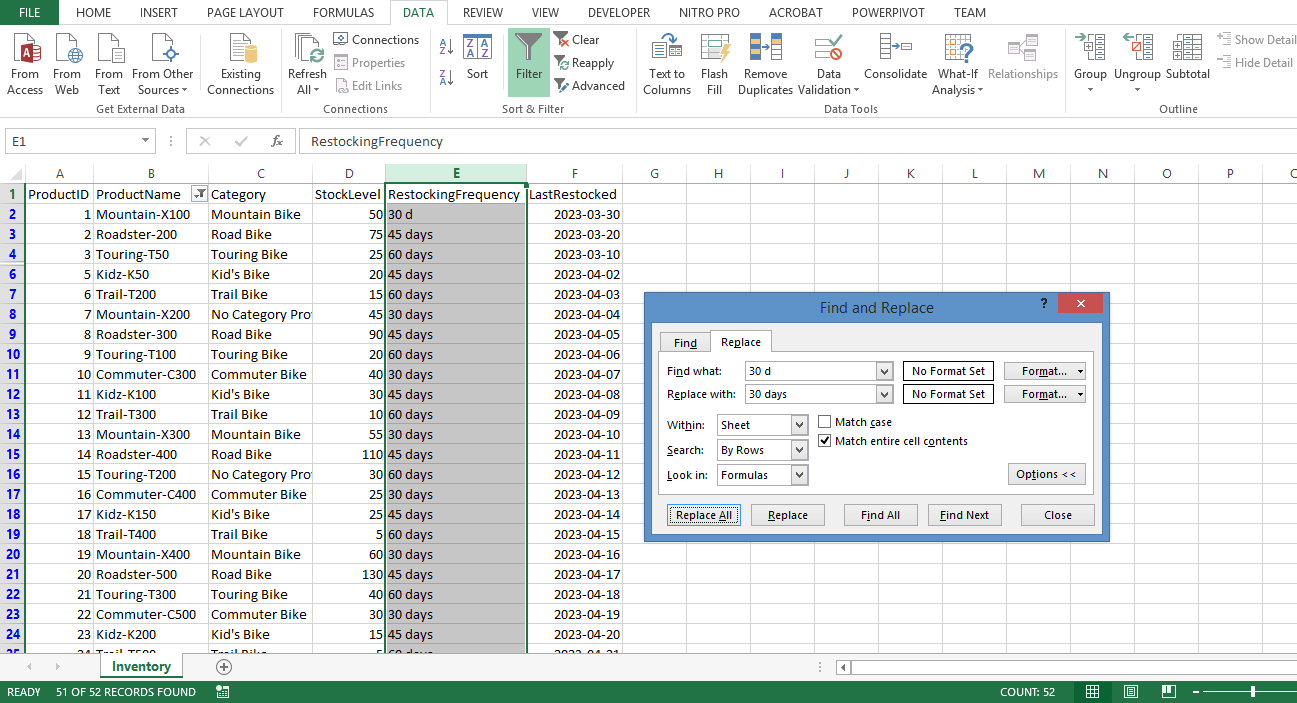


**Step 3: Standardizing RestockingFrequency**

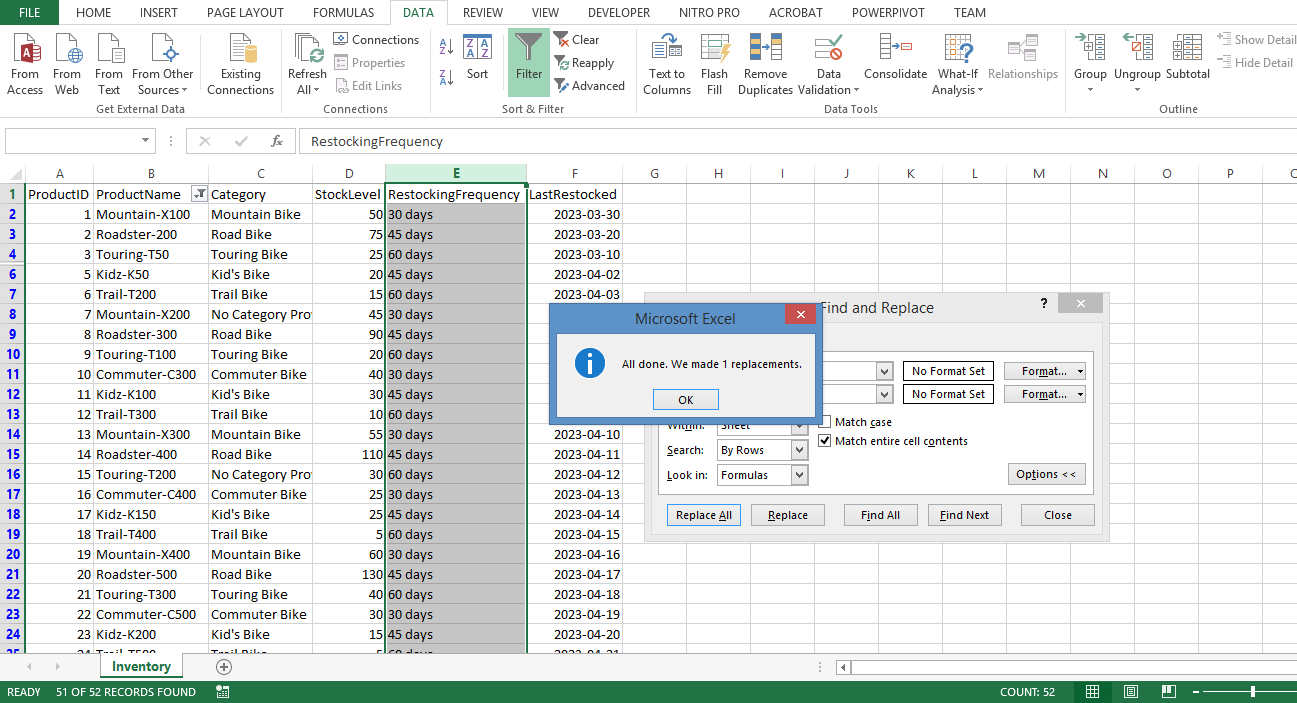
1. Inconsistencies in **RestockingFrequency** can also hinder accurate analysis. You found the **RestockingFrequency** column and selected it.
2. You pressed **CTRL + H** to open the **Find and Replace** dialog.



1. In the **Find what** field, you typed **30 d**, and replaced it with **30 days**.



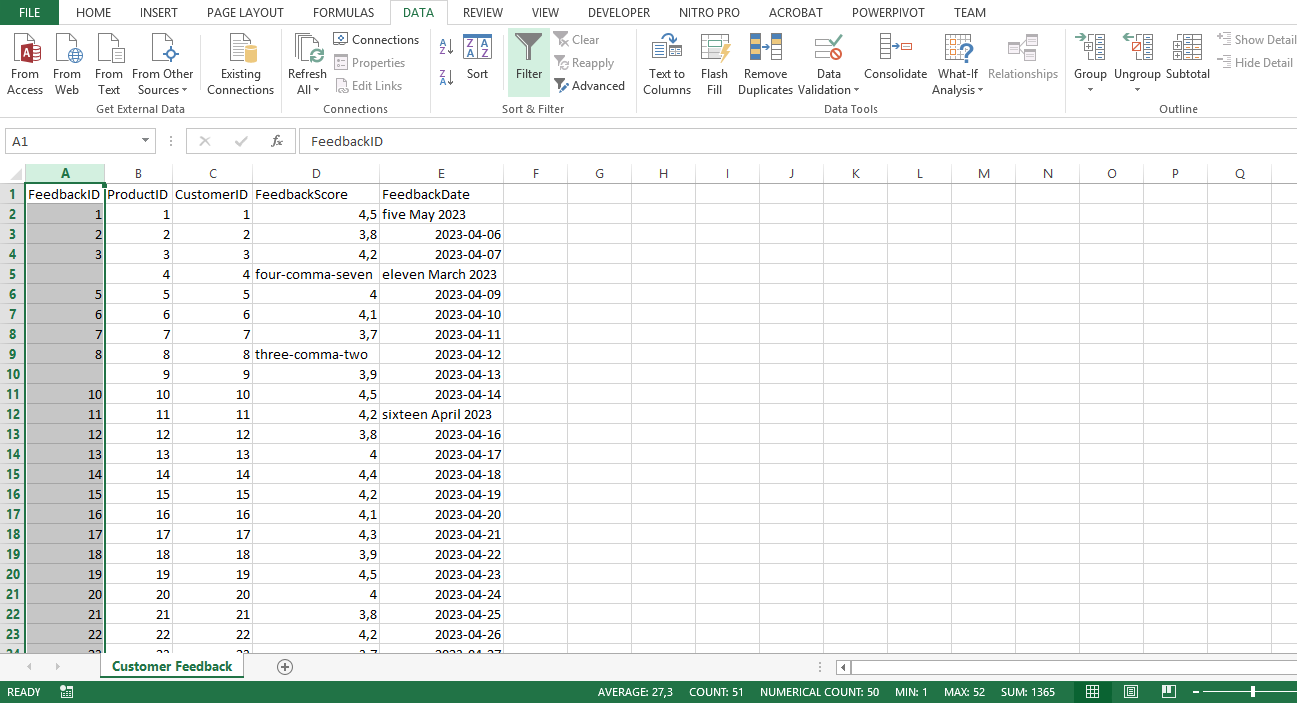
4. Here you expanded the **Options** menu, and selected **Match entire cell contents,** and selected **Replace All**. This step helped to maintain uniformity across data entries. With the *Inventory data* clean and standardized, when the data is loaded into Power BI you were better able to answer the question: 'What is the stock level for each bike category?'



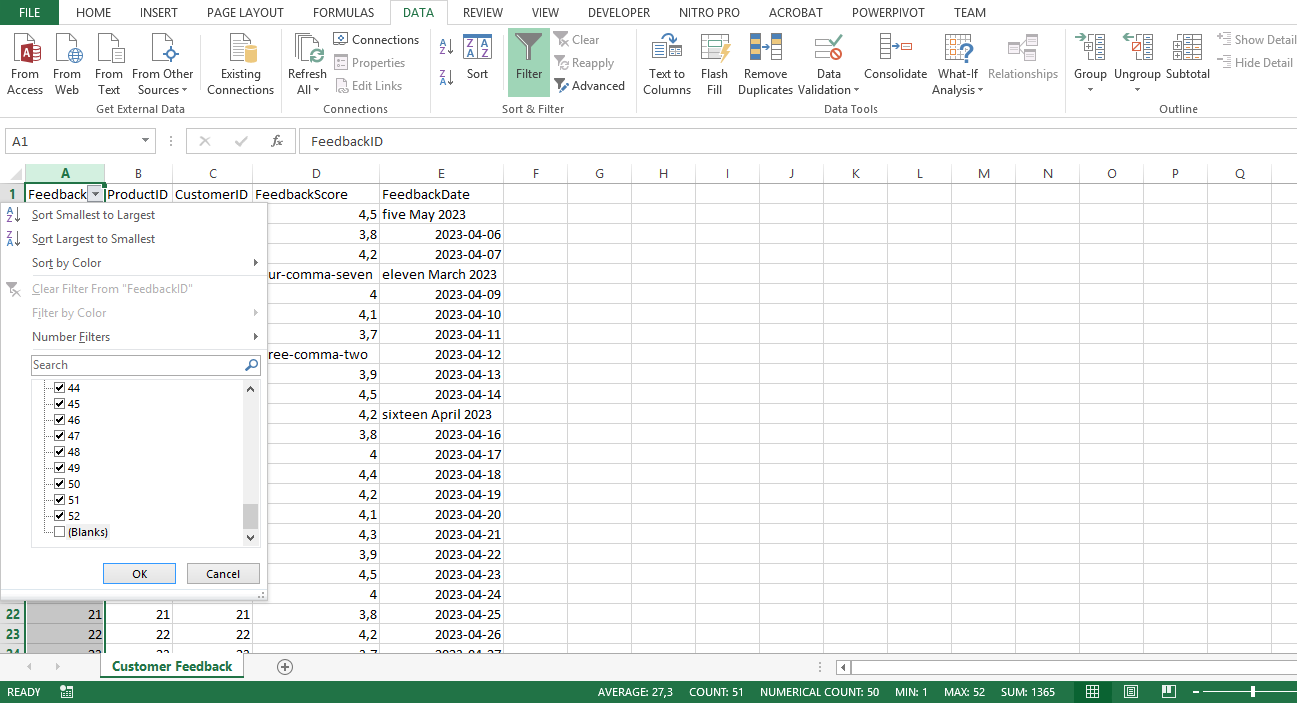
**Customer Feedback Dataset: What is the average feedback score for each product?**

**Step 1: Addressing Missing FeedbackID**

1. You began by filtering out any rows with missing **FeedbackIDs** in the *Customer Feedback* data. Similar to missing **TransactionIDs** and **ProductNames**, missing **FeedbackIDs** can render the corresponding data points anonymous and potentially unreliable. You selected the **FeedbackID** column.

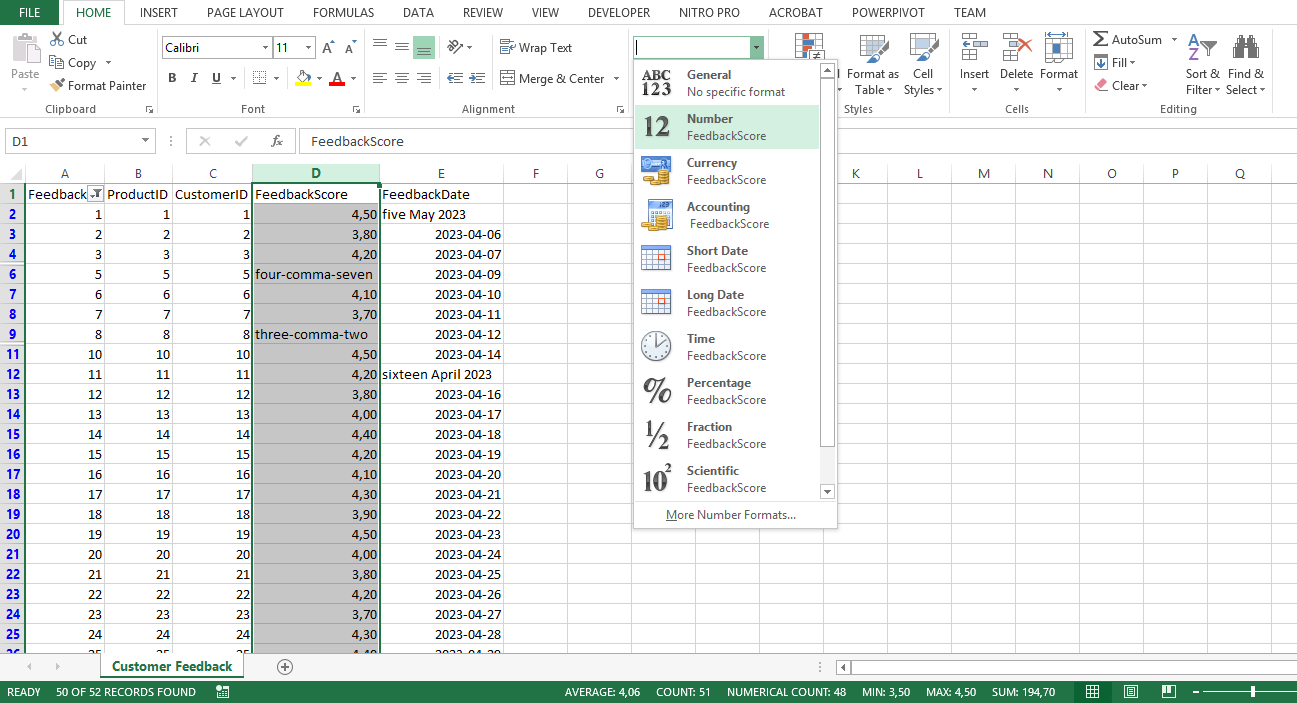


1. On the **Data** tab, you selected **Filter**.
2. Selecting the drop-down arrow on the **FeedbackID** column you deselected **Blanks** and selected **OK**.

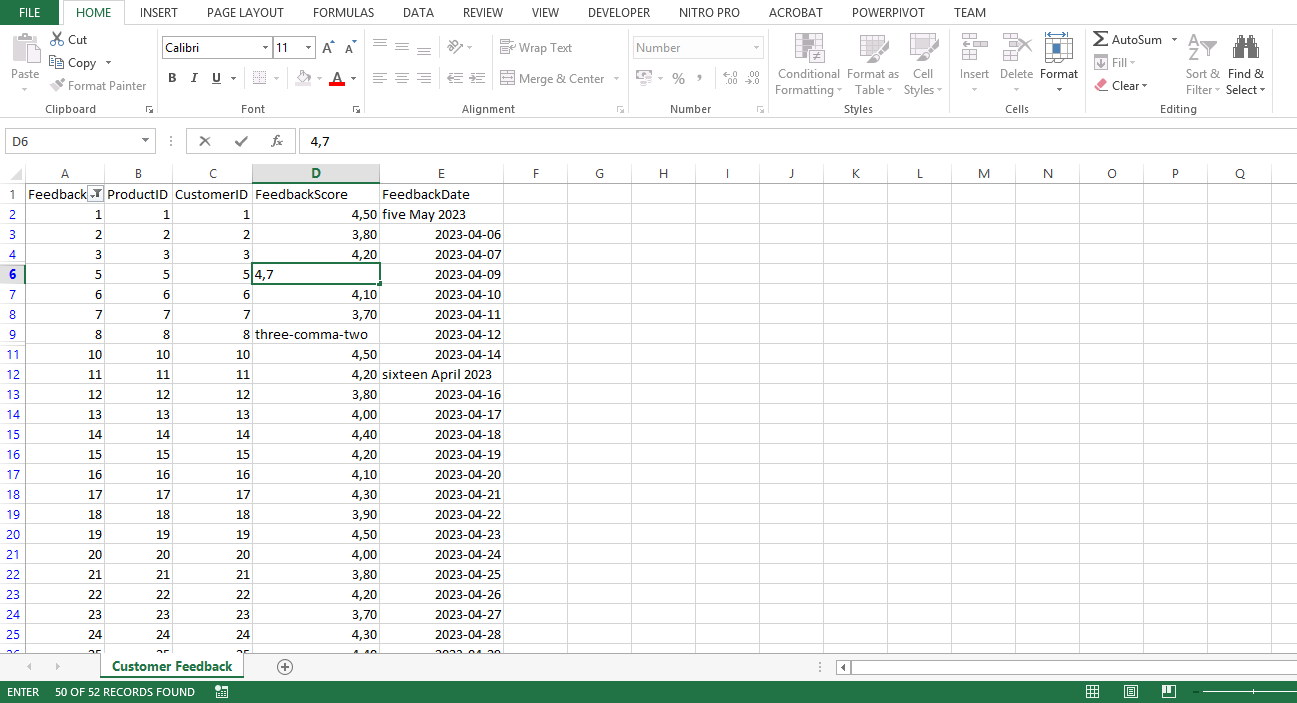


**Step 2: Dealing with inconsistent FeedbackScore**

1. You dealt with inconsistencies in the **FeedbackScore** column by setting the data type to **Number**, ensuring all scores were treated as numerical values.

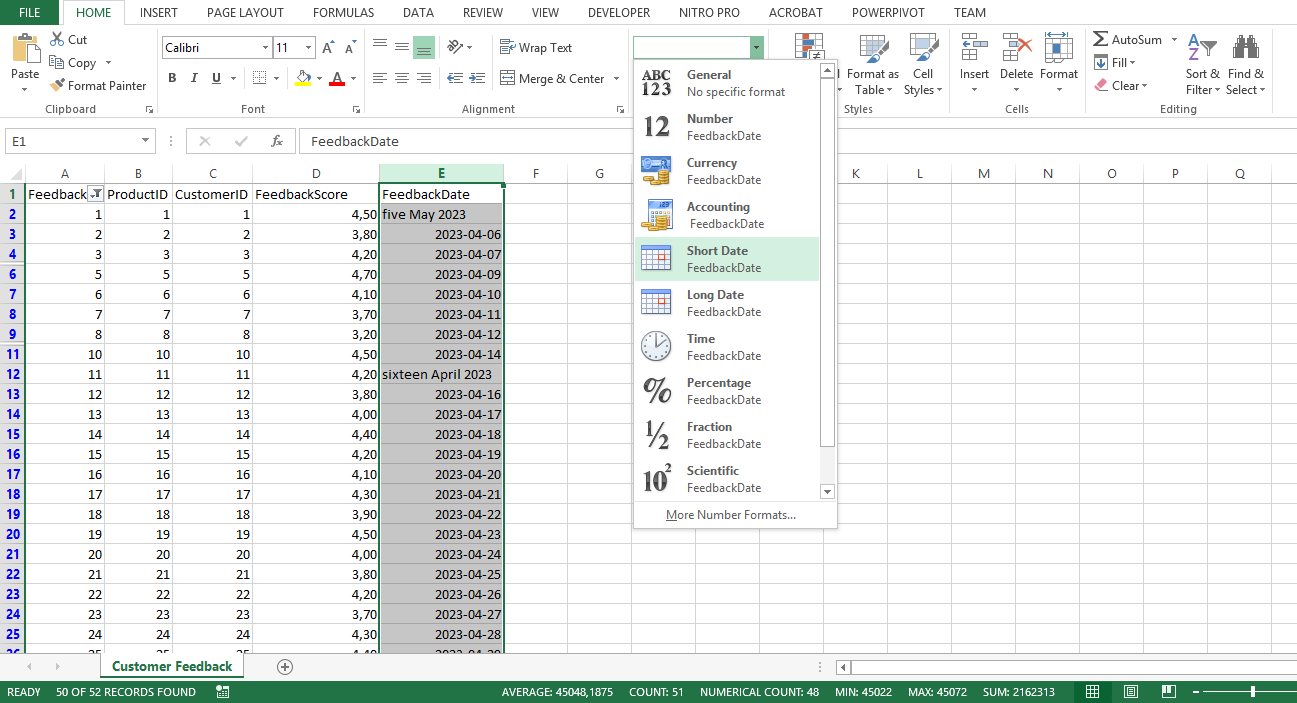


1. You manually edited the content of the column to identify and replace non-numeric values. To do this, you double-clicked on the cell with the non-numeric value, deleted the existing value and typed the appropriate numeric value.

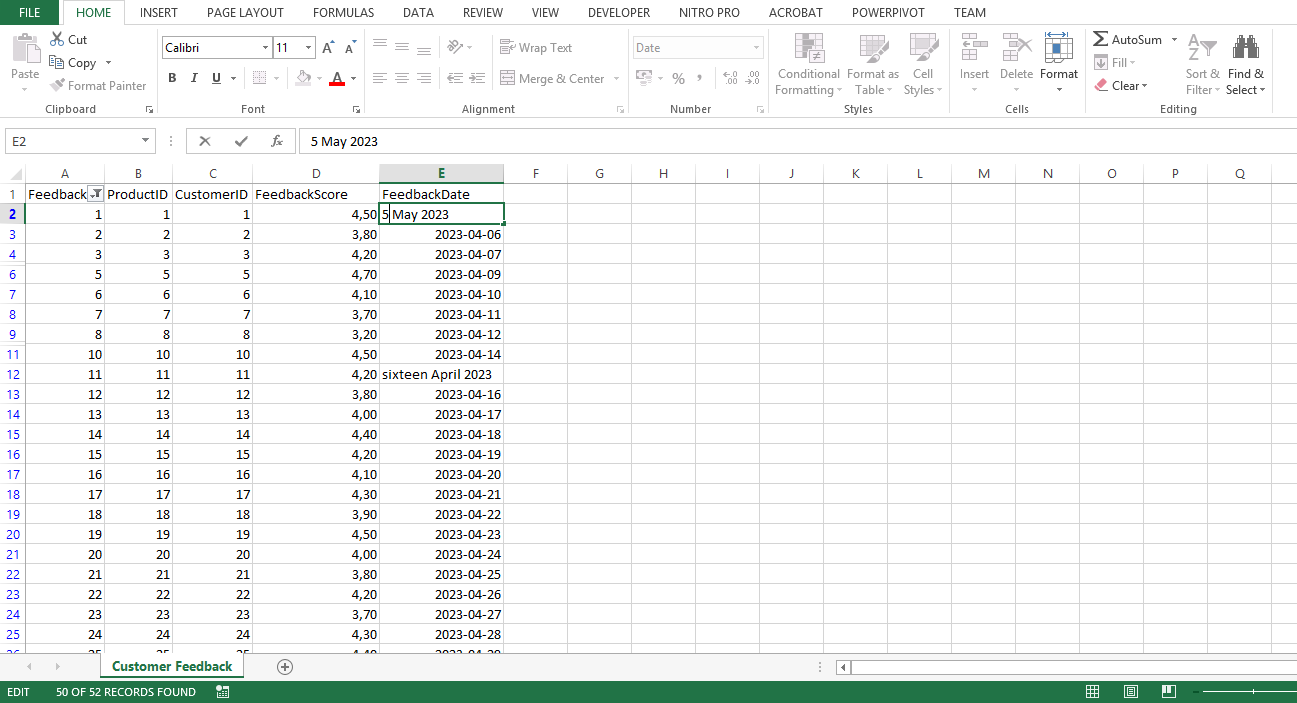


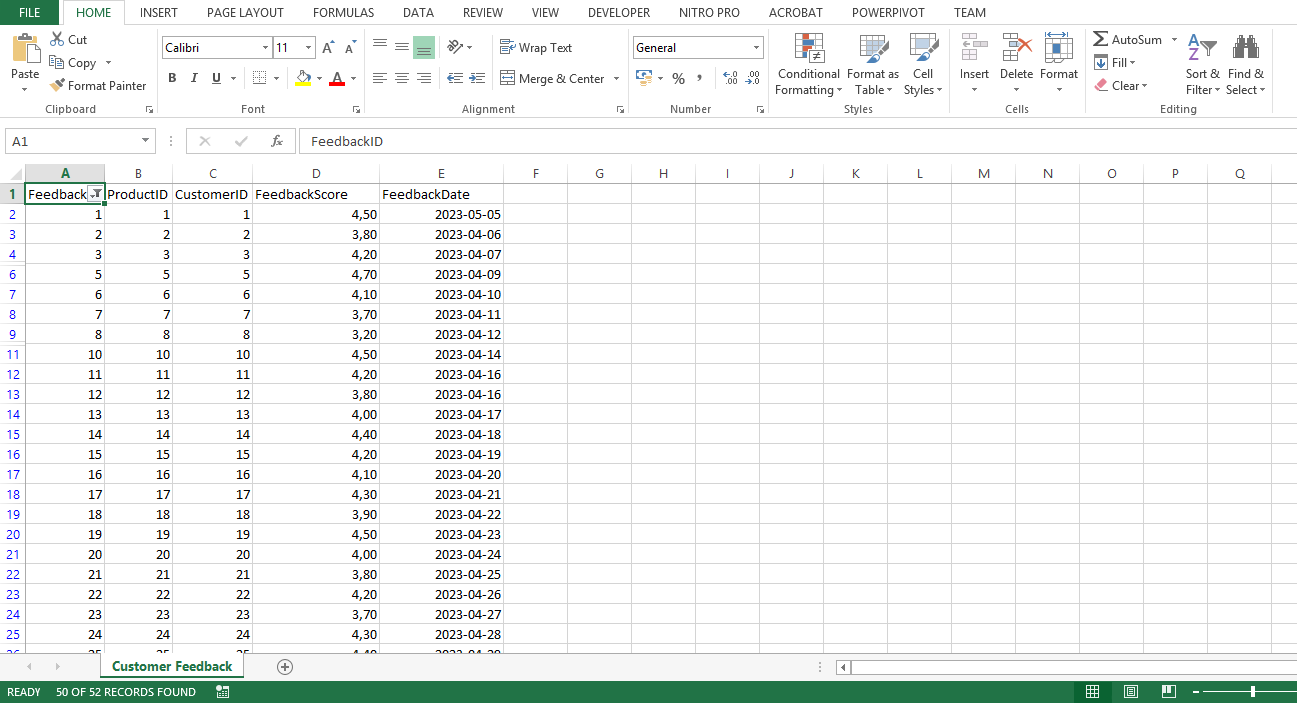
**Step 3: Invalid dates in the FeedbackDate column**

1. Next, you addressed erroneous dates in the **FeedbackDate** column.
2. You selected the **FeedbackDate** column and set the data type to **Date** on the ribbon, which forced Excel to interpret all entries as dates.



1. You then manually identified and replaced non-date values in the column. To do this, you double-clicked on the cell with the non-numeric value, deleted the existing value and typed the appropriate numeric value. Evaluation of the *Customer Feedback* data was then complete, ready for future analysis in Power BI.





**Conclusion**

By following these detailed, step-by-step instructions, you have demonstrated proficiency in evaluating data with Microsoft Excel, and preparing the data for further transformation and analysis in Power BI. As you continue working with different datasets and analysis questions, you will become more comfortable with the process and better equipped to draw valuable insights from your data. Keep exploring and pushing your data analysis skills to new heights!