# Activity

# Importing web data with Power Query

qa.com

# 1 Introduction

In this activity, we will use the Power Query feature of Microsoft Excel to download some historical share price data from a web source and display these data in an Excel table.

We will also see how to edit, refresh, and duplicate Queries.

## 1.1 Prerequisites

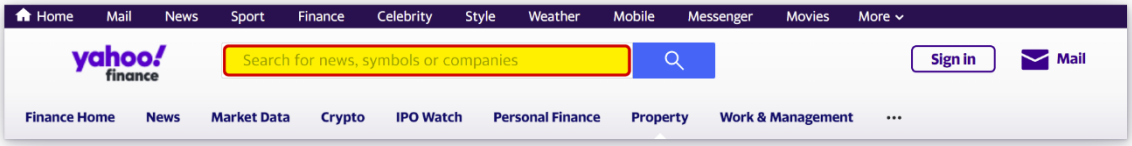
Familiarity with Microsoft Excel is helpful, but not essential.

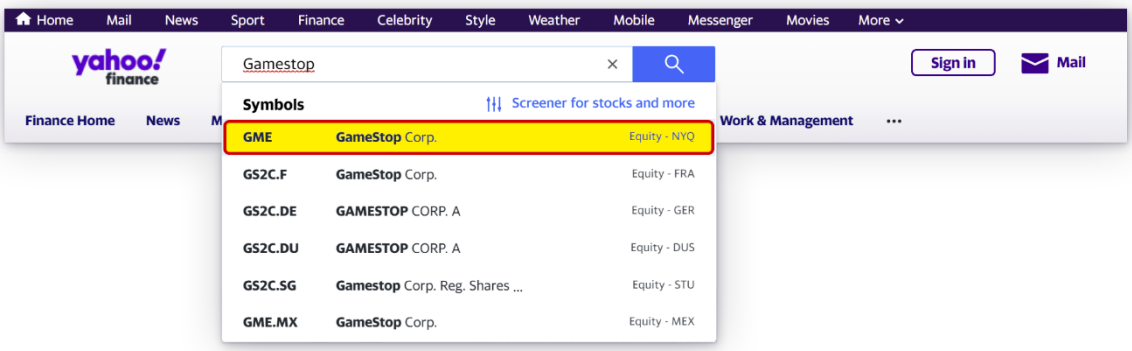
# 2 Create a Query

In this step, we will create a Query that downloads a CSV file from the web, parses and tidies the file contents, and displays the output in a table on the worksheet.

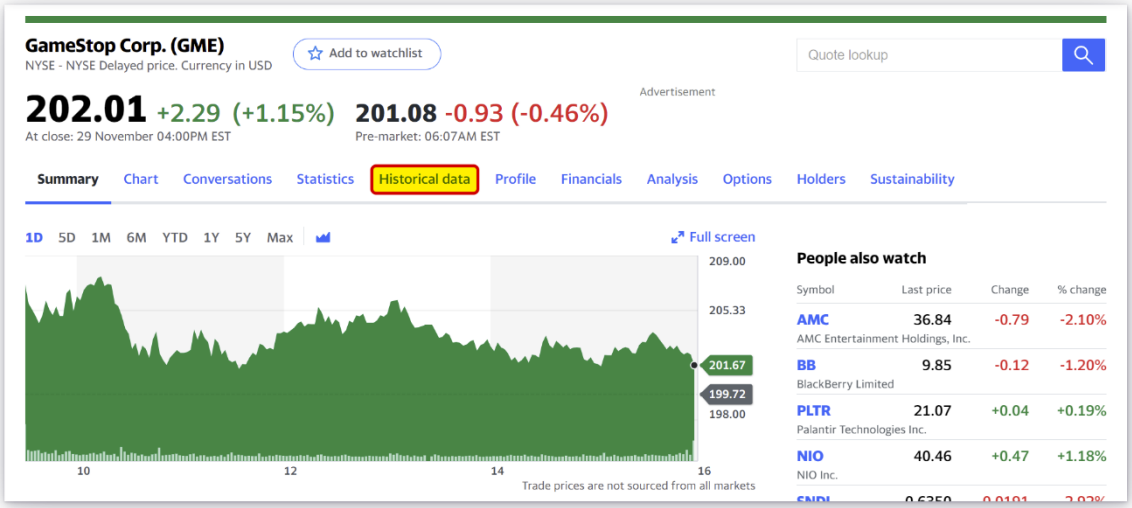
## 2.1 Download share price data

1. Navigate to [uk.finance.yahoo.com](uk.finance.yahoo.com)
2. Search for the name or abbreviation of a share: for example, GameStop Corp. (GME).



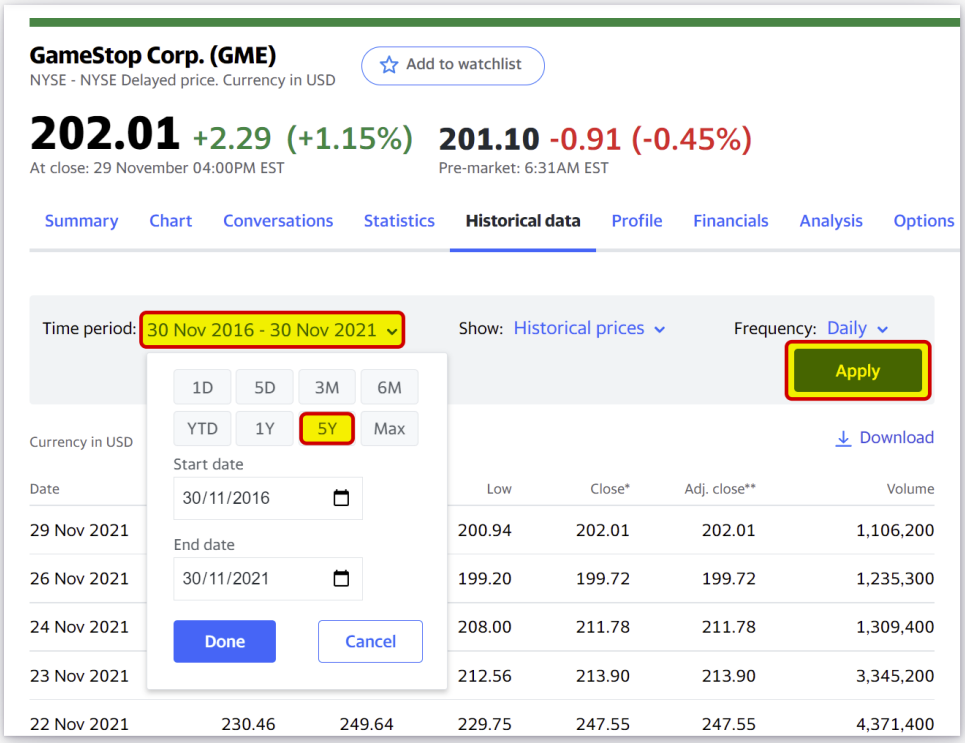



3. Open the **Historical data** tab.

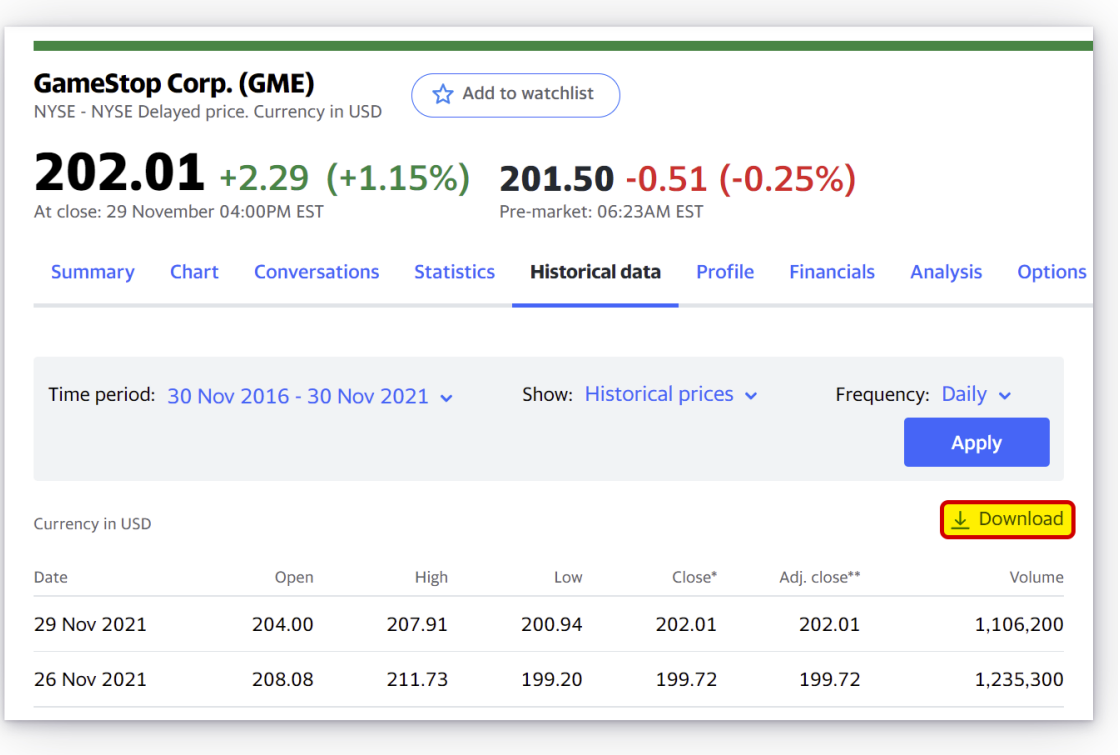



4. Filter the time period to **5Y**. This is to use the data from the last 5 years. Click **Apply**.

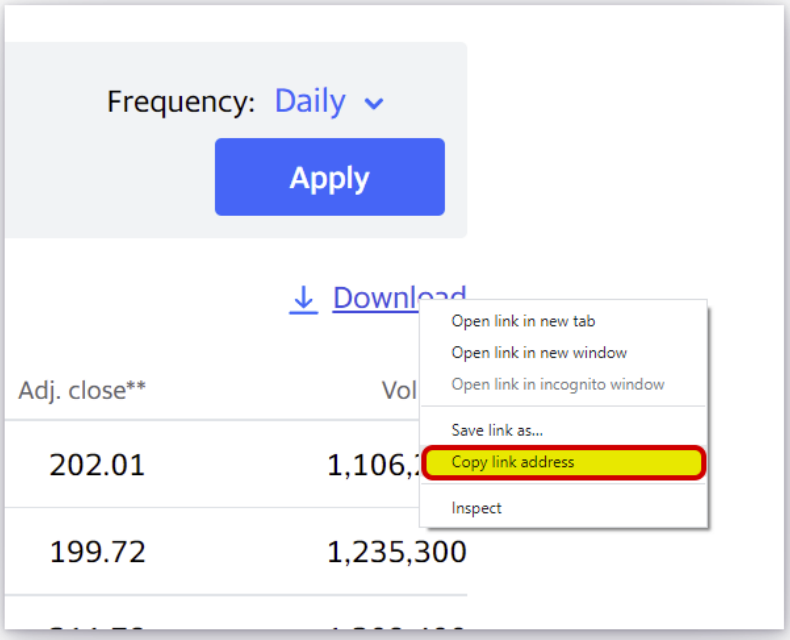
**NOTE:** The dates on the below screenshot will differ to your current dates. You should have the present dates when you are working on this activity up until the last 5 years.



5. Verify that **Time period** now displays the period you selected. Right-click the **↓ Download** button.

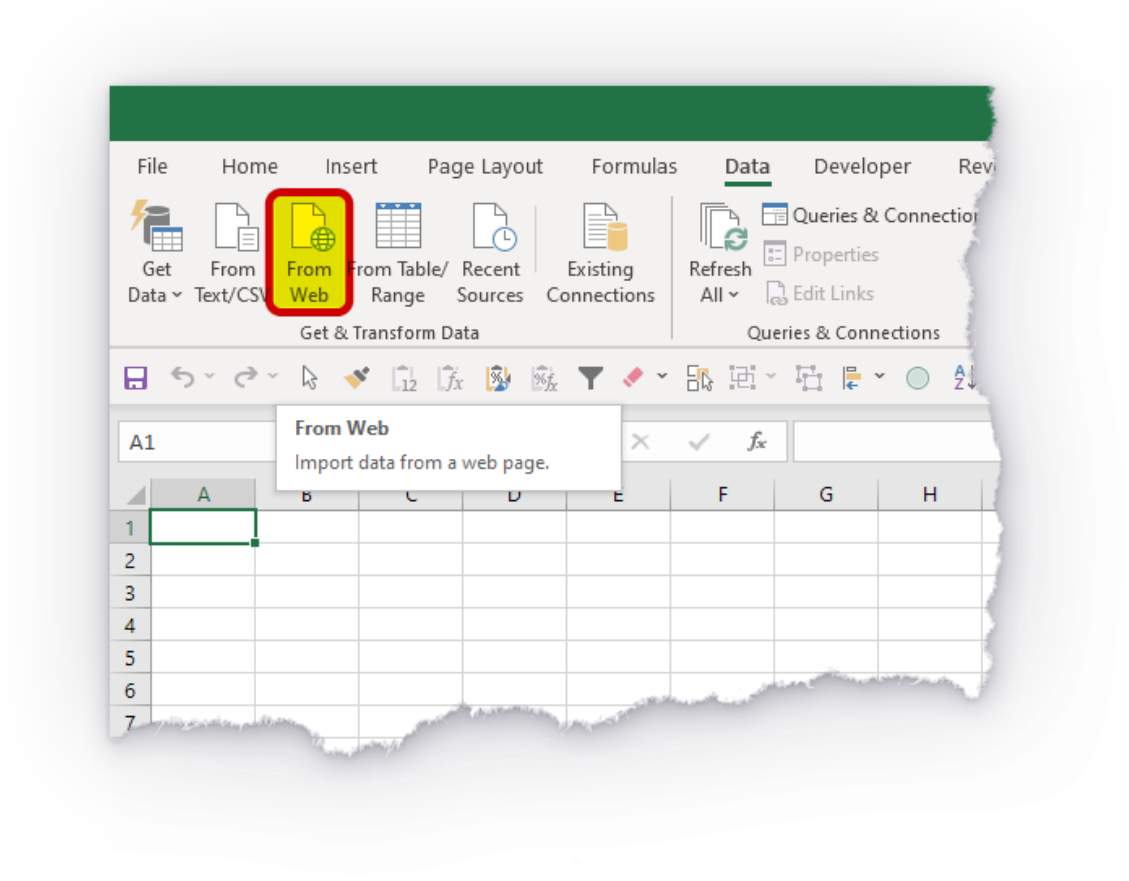



6. Click **Copy link address**.

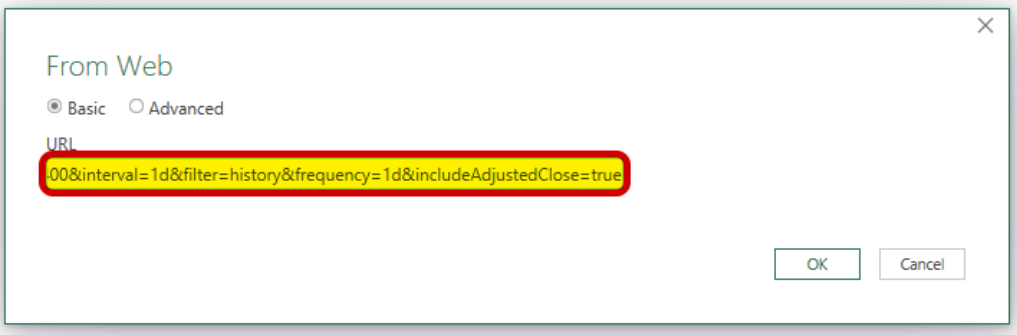


## 2.2 Create a new Query

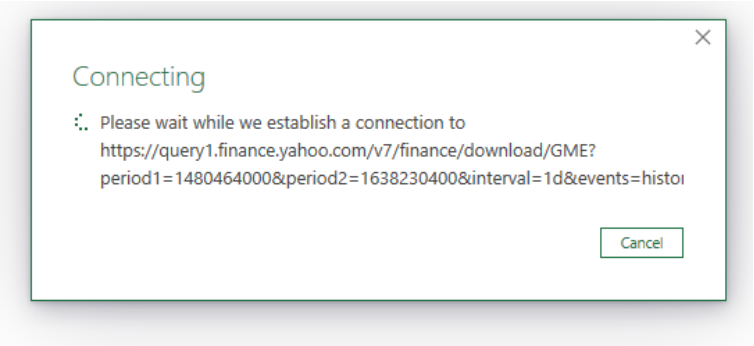
7. In a new Excel workbook, navigate to **Ribbon** > **Data** tab > **Get & Transform Data** group > **From Web**.




8. In the dialog that opens, paste (**Ctrl + V**) the link you copied earlier into the **URL** field. The URL should start with 'query1.finance.yahoo.com...'. Leave the **Basic** radio button selected. Click **OK**.

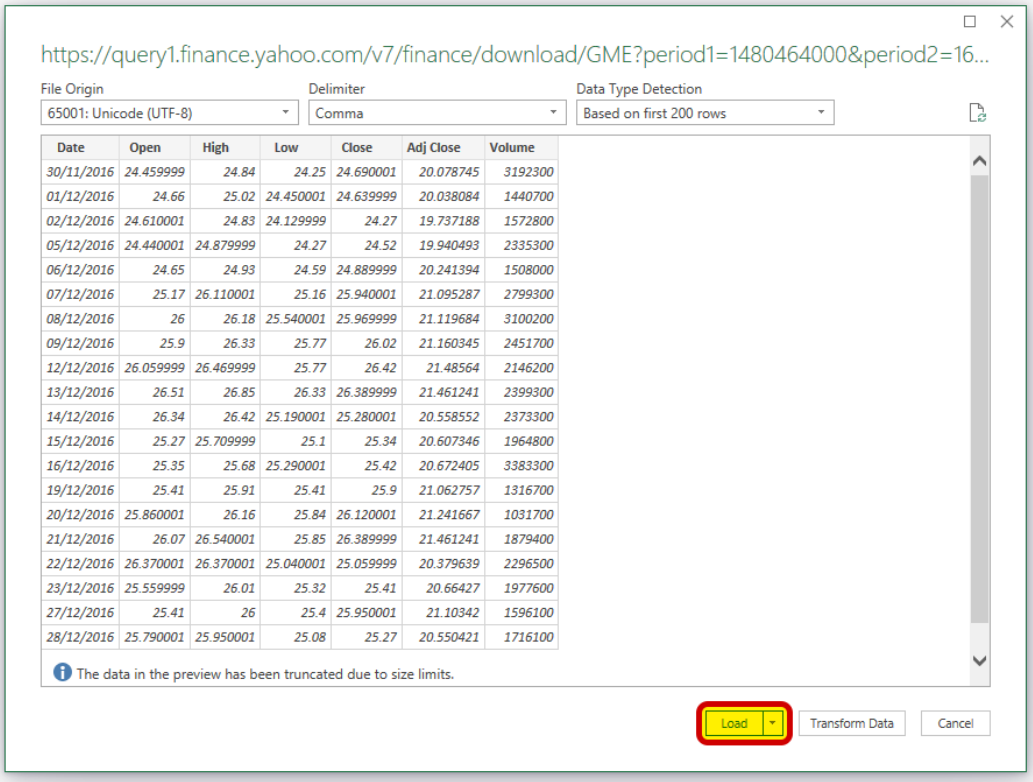


9. The **Connecting** dialog will show for a few seconds.



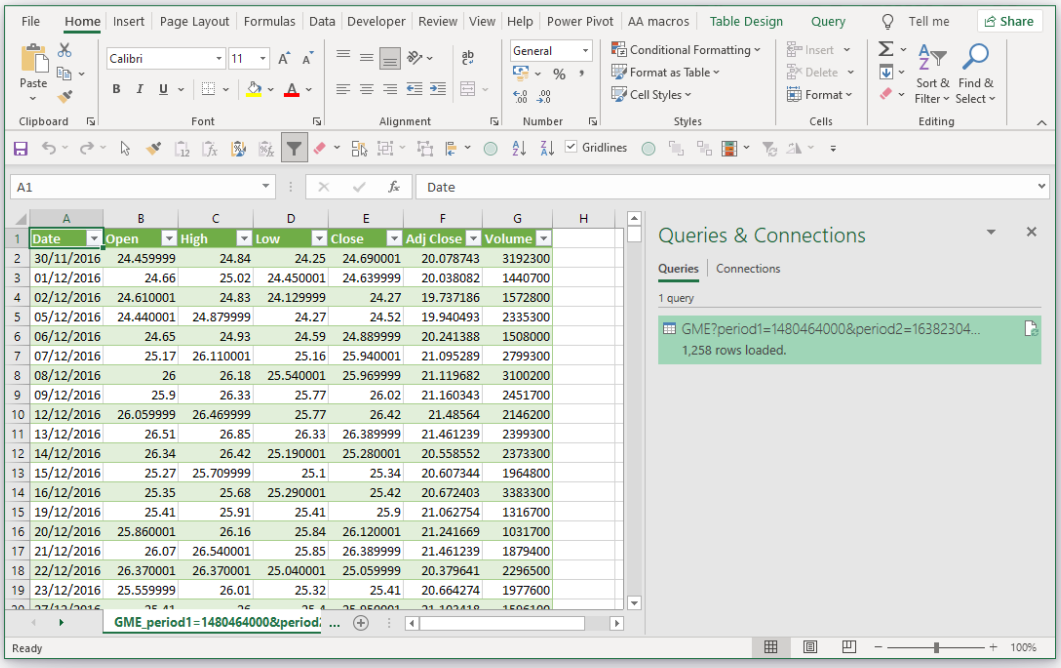

10. A preview of the table appears. Excel detects that the data source is a CSV file, and it makes reasonable guesses for the character encoding (UTF-8) and value separator (comma).

    The data in the table look good (no corrupted characters, misaligned columns etc.), so let's click **Load**.




(Note: if we were to instead click **Transform Data**, this would bring up the Power Query Editor window. We will explore that feature later.)

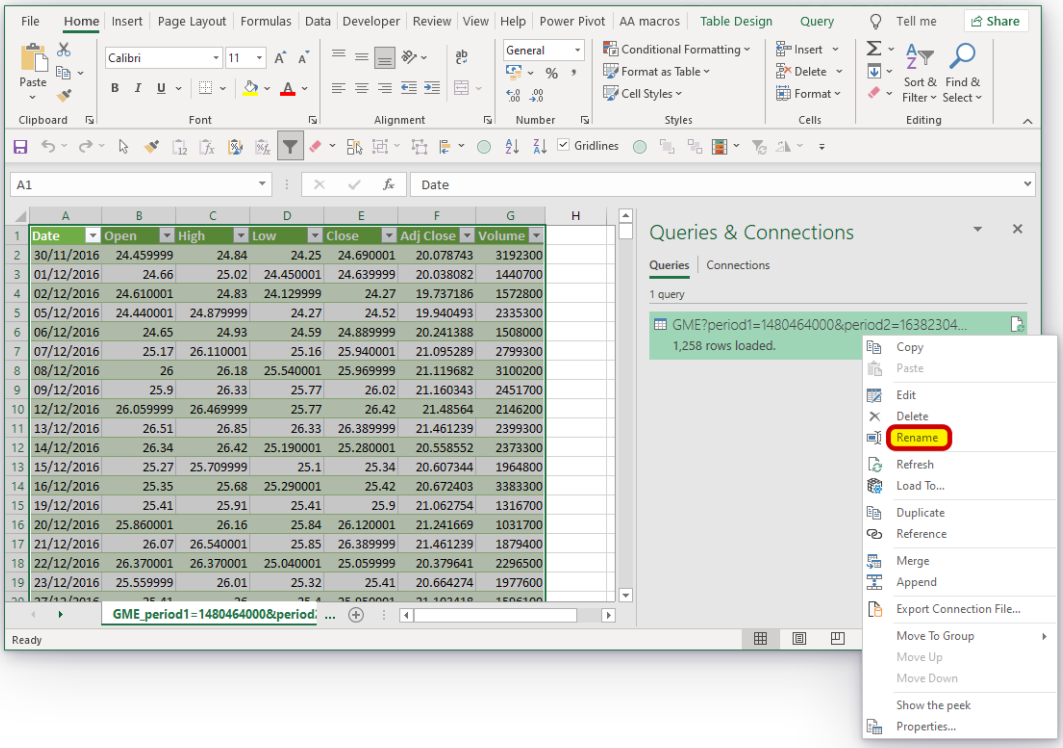
11. The data source is loaded as a table on a new worksheet. We also see that a panel called **Queries & Connections** has appeared. This panel displays the name of the Query associated with the data in the table. Later, we will see how we can edit and refresh this Query.



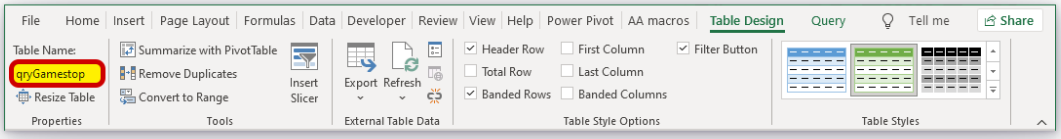

| Date | Open | High | Low | Close | Adj Close | Volume |
|---|---|---|---|---|---|---|
| 30/11/2016 | 24.459999 | 24.84 | 24.25 | 24.690001 | 20.078743 | 3192300 |
| 01/12/2016 | 24.66 | 25.02 | 24.450001 | 24.639999 | 20.038082 | 1440700 |
| 02/12/2016 | 24.610001 | 24.83 | 24.129999 | 24.27 | 19.737186 | 1572800 |
| 05/12/2016 | 24.440001 | 24.879999 | 24.27 | 24.52 | 19.940493 | 2335300 |
| 06/12/2016 | 24.65 | 24.93 | 24.59 | 24.889999 | 20.241388 | 1508000 |
| 07/12/2016 | 25.17 | 26.110001 | 25.16 | 25.940001 | 21.095289 | 2799300 |
| 08/12/2016 | 26 | 26.18 | 25.540001 | 25.969999 | 21.119682 | 3100200 |
| 09/12/2016 | 25.9 | 26.33 | 25.77 | 26.02 | 21.160343 | 2451700 |
| 12/12/2016 | 26.059999 | 26.469999 | 25.77 | 26.42 | 21.48564 | 2146200 |
| 13/12/2016 | 26.51 | 26.85 | 26.33 | 26.389999 | 21.461239 | 2399300 |
| 14/12/2016 | 26.34 | 26.42 | 25.190001 | 25.280001 | 20.558552 | 2373300 |
| 15/12/2016 | 25.27 | 25.709999 | 25.1 | 25.34 | 20.607344 | 1964800 |
| 16/12/2016 | 25.35 | 25.68 | 25.290001 | 25.42 | 20.672403 | 3383300 |
| 19/12/2016 | 25.41 | 25.91 | 25.41 | 25.9 | 21.062754 | 1316700 |
| 20/12/2016 | 25.860001 | 26.16 | 25.84 | 26.120001 | 21.241669 | 1031700 |
| 21/12/2016 | 26.07 | 26.540001 | 25.85 | 26.389999 | 21.461239 | 1879400 |
| 22/12/2016 | 26.370001 | 26.370001 | 25.040001 | 25.059999 | 20.379641 | 2296500 |
| 23/12/2016 | 25.559999 | 26.01 | 25.32 | 25.41 | 20.664274 | 1977600 |

12. The default name of our Query is long and unreadable. To change it, right-click the name of the Query in the **Queries & Connections** panel and click **Rename**. Give the Query a short name such as ‘qryGamestop’. (The ‘qry’ prefix helps remind us that this is a Query.)

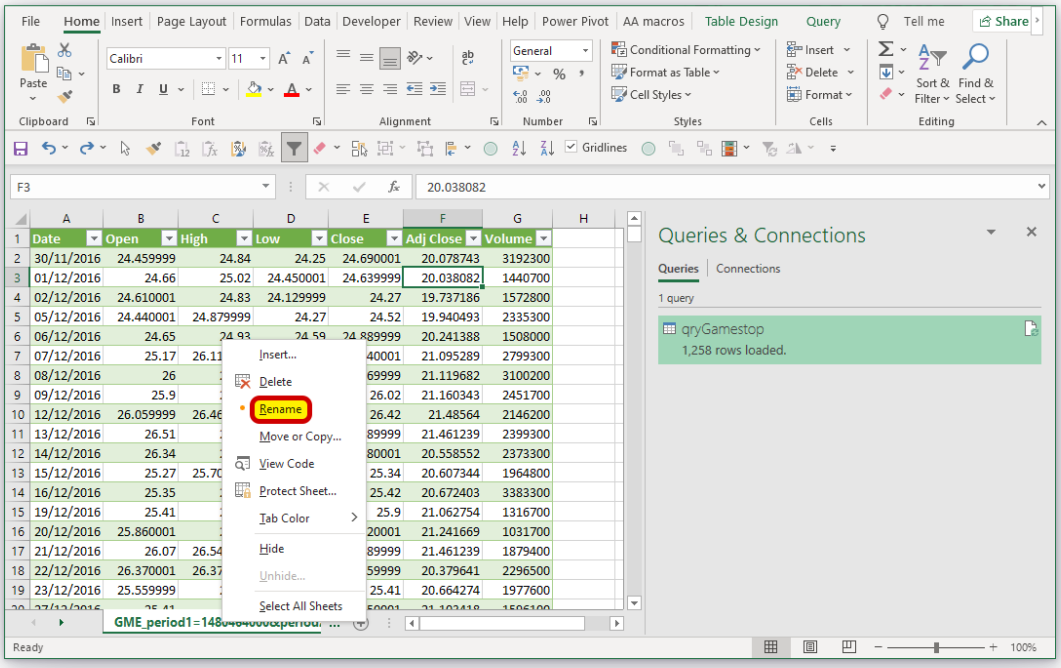
From this point on, the share price Query will be referred to as **qryGamestop**. The name of your Query might be different.




13. If we navigate to **Ribbon** > **Table Design** tab, we can see that the table adopts the same name as its associated Query.



14. Let’s also rename the worksheet to ‘qryGamestop’.

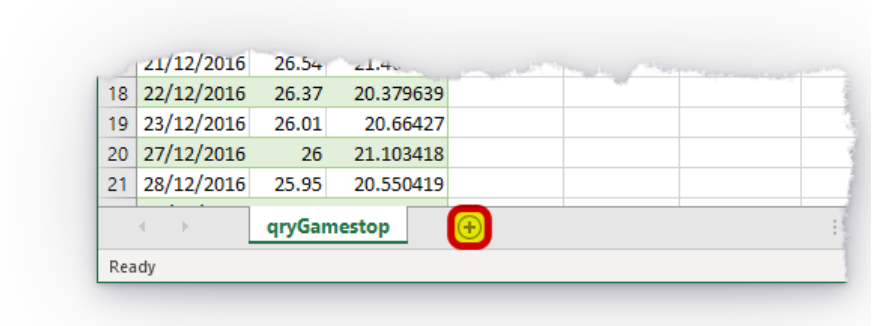



## 2.3 Plot the Query results

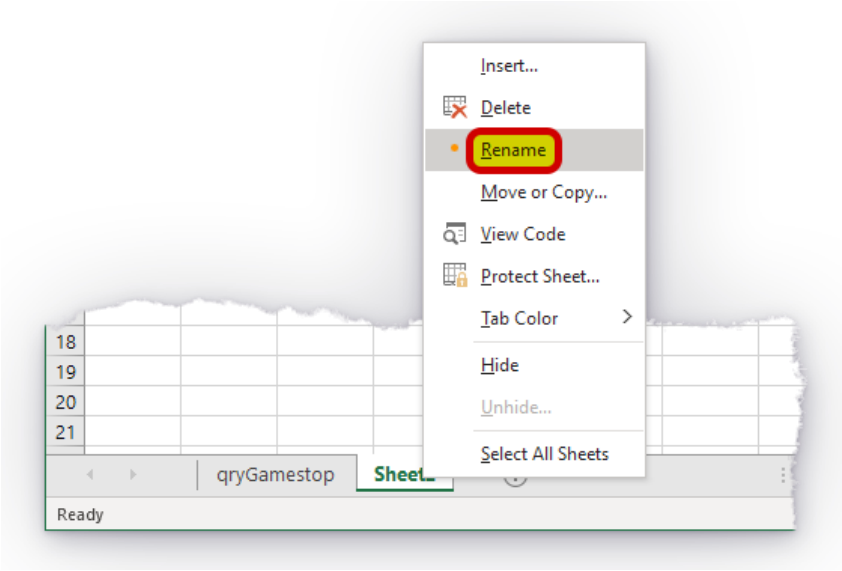
On a separate worksheet within the same workbook, we will create two scatter charts:

- **Chart 1:** Adj. Close vs. Date.
- **Chart 2:** Volume vs. Date.

15. First, create a new worksheet by clicking the ⊕ symbol on the worksheet tab bar at the bottom of the Excel window.



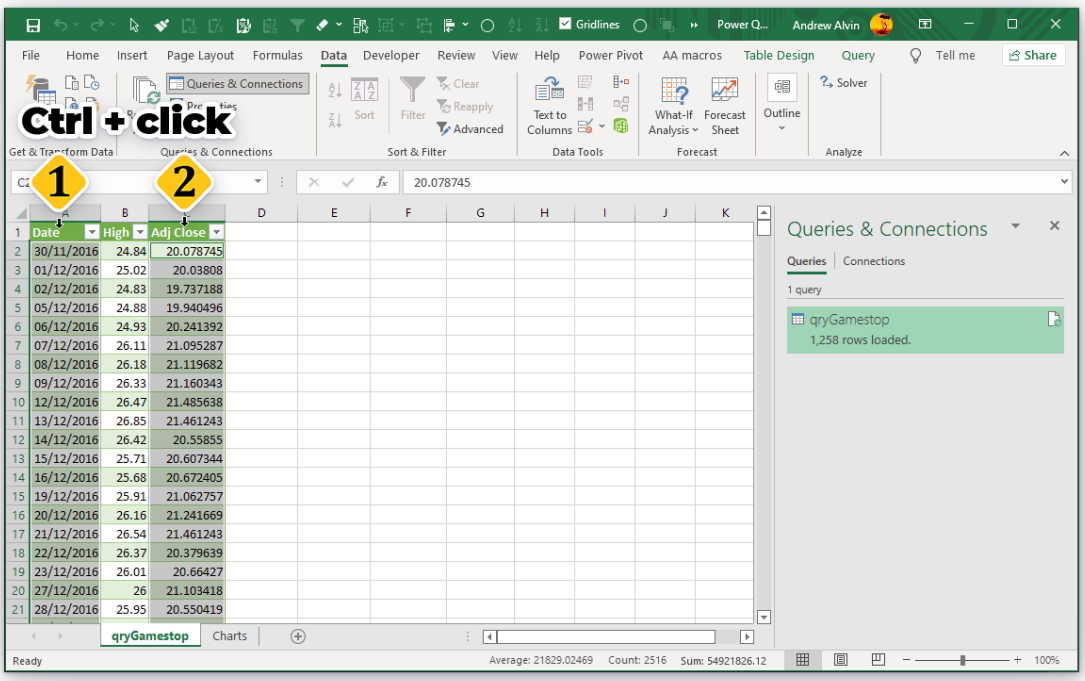
16. Name the worksheet ‘Charts’.



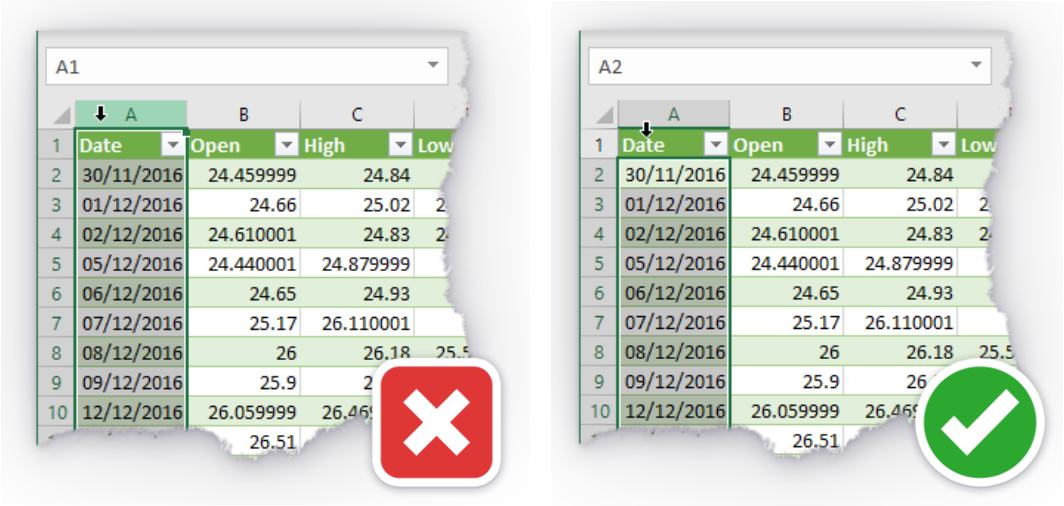

17. There are several ways of creating a chart in Excel; we will walk through one of the simplest.

    First, navigate to the **qryGamestop** worksheet. Scroll to the top of the sheet, so that you can see the header row of the **qryGamestop** table. Then, select the columns you want to plot by first clicking once on the first column, **‘Date’** as shown in the following screenshots, then hold the **Ctrl** key and click **‘Adj Close’** column.

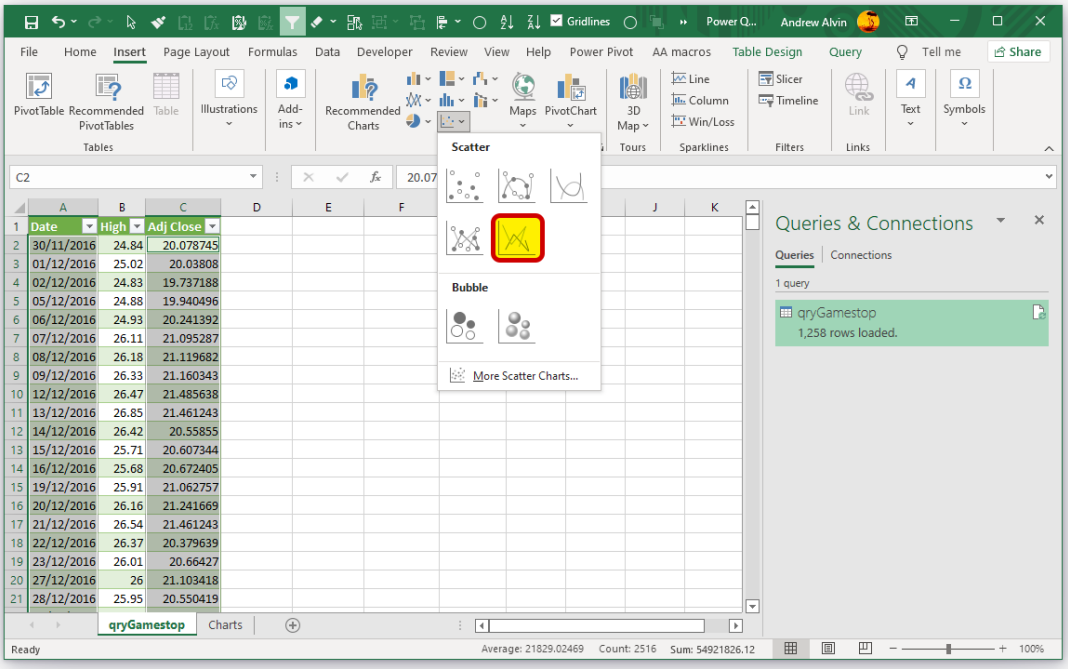
    **Note:** you are looking for the mouse pointer to turn into a down arrow (⬇). This occurs when the mouse is positioned near the top of the header cell as shown below.



⚠ When you are selecting the source data for the charts, ensure that you select just the table column (approx. 1,000 rows) and not the entire worksheet column (> 1 million rows). By selecting just the table column, we create a dynamic reference to that column which will move, grow, and shrink with the table.



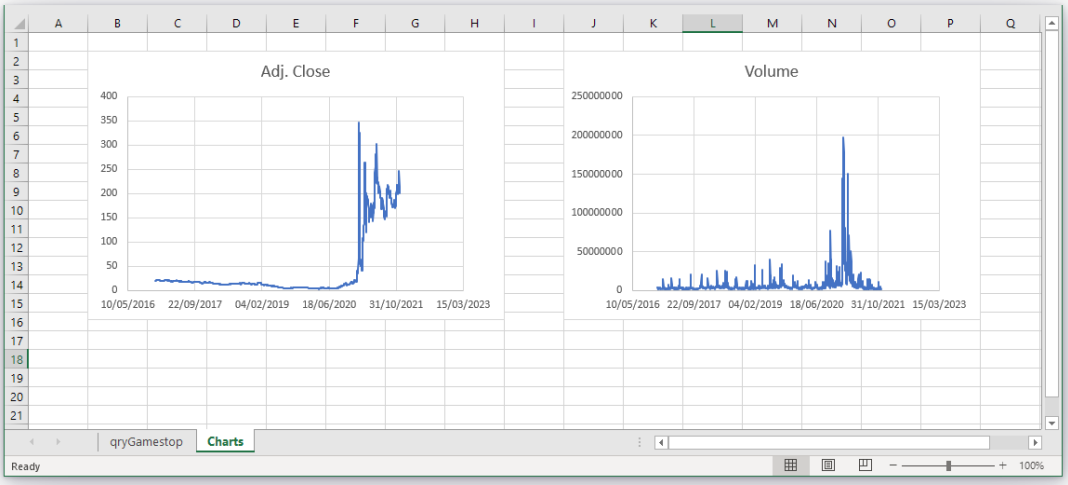

18. Navigate to **Ribbon** > **Insert** > **Charts** group > **Insert Scatter or Bubble Chart** button > **Scatter with Straight Lines**.




19. Select the resulting chart, then cut it (Ctrl + X), navigate to the Charts worksheet, and paste (Ctrl + V) the chart there.

Repeat steps 17–19, this time plotting the ‘Date’ and ‘Volume’ columns. Remember to click on the ‘Date’ column first then hold the **Ctrl** key and click the **‘Volume’** column.

20. Your charts should look similar to those below.



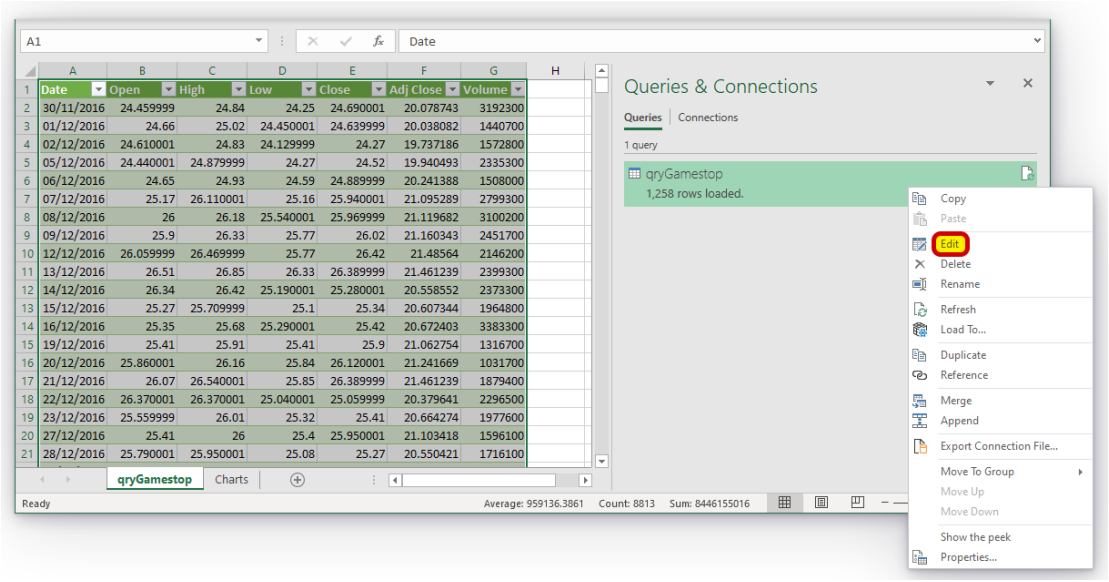

# 3 Edit the Query

## 3.1 Open the Power Query Editor

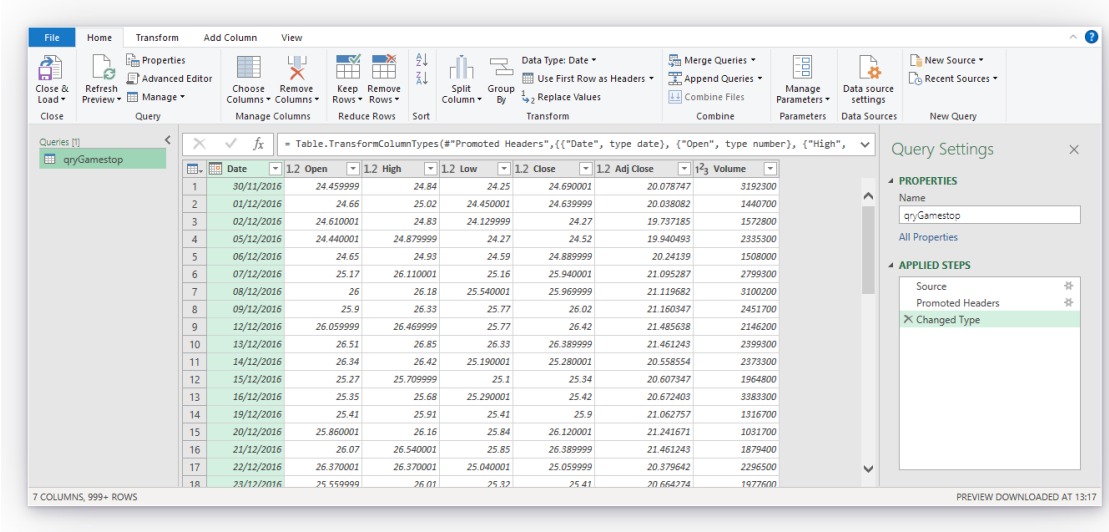
21. We will now edit our Query by right clicking the Query name in the **Queries & Connections** panel and clicking **Edit**. Alternatively, you can double-click the Query name.

    (If the **Queries & Connections** panel is closed, open it by clicking **Ribbon** > **Data** tab > **Queries & Connections** group > **Queries &**

**Connections**.)

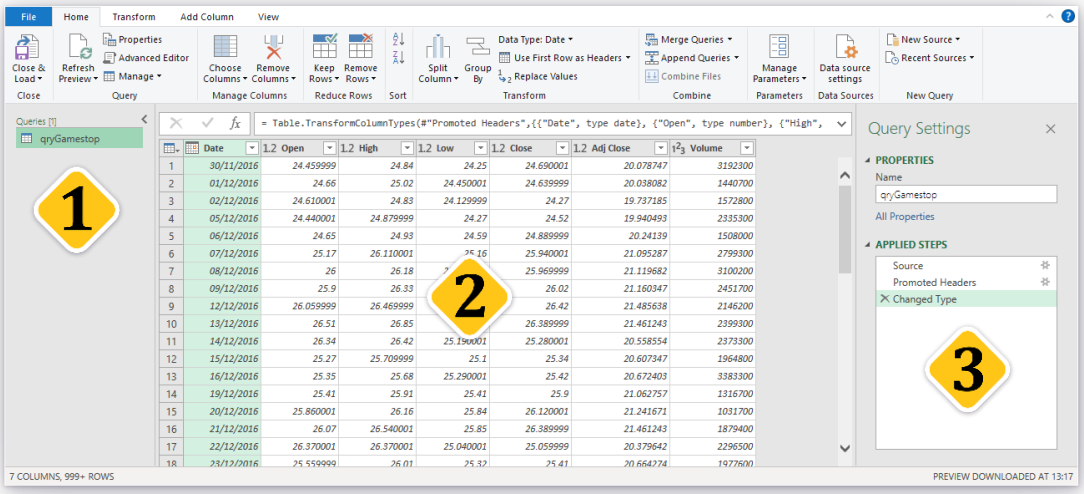



This opens the Power Query Editor window.




Power Query is a data transformation tool built into Microsoft Excel and Microsoft Power BI. With Power Query, we can create pipelines for retrieving, cleaning, and transforming data. This type of processing is often referred to as ETL (Extract, Transform, Load).

22. An overview of the Power Query interface.




*1)* The **Queries** panel: a navigable list of all Queries in this workbook.

*2)* Data preview. Note: the table data cannot be modified by manually typing into the cells. In Power Query, the data set is modified by applying transformations to rows and columns.

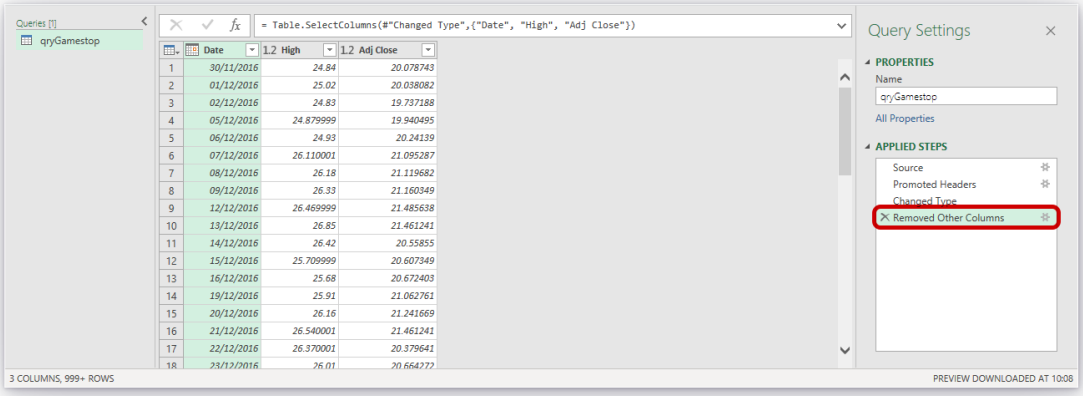
*3)* The **Applied Steps** list: an ordered list of data transformations that have been applied to the source data.

23. In the **Applied Steps** list, we can see that three transformation steps have already been populated for us automatically: 'Source', 'Promoted Headers', and 'Changed Type'. This was a result of using the **Get & Transform Data** > **From Web** wizard. Alternatively, we could have created these steps manually within the Power Query Editor.

24. By clicking a step in the **Applied Steps** list, we can 'rewind' or 'fast-forward' to that step. Try clicking on each of the existing steps in turn to see how they transform the data. When you are done, click on the final step, 'Changed Type', before continuing.

25. We will now add a new step to the **Applied Steps** list. Navigate to **Ribbon** > **Home** tab > **Manage Columns** group > **Choose Columns**. Select the following columns: ‘Date’, ‘High’ and ‘Adj Close’. Click **OK**.

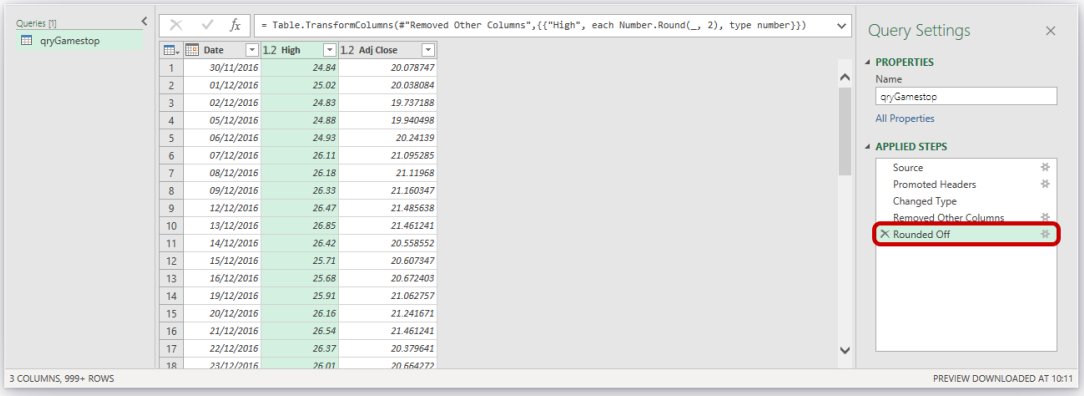

We see that all columns have been hidden except for the three columns we selected in the **Choose Columns** dialog. Also, a new step has been appended to the **Applied Steps** list: ‘Removed Other Columns’.




Note: in Power Query, all data transformations (applied steps) are non-destructive. We can roll back to a previous step by clicking on it in the **Applied Steps** list. We can even edit a previous step by double-clicking it. Power Query will warn us if our changes are likely to have destructive knock-on effects on subsequent steps.

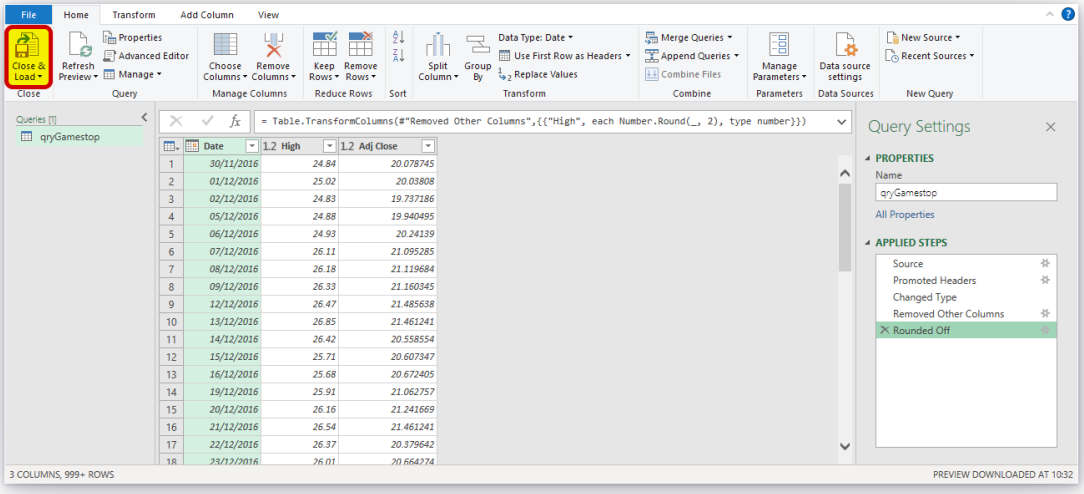
26. In the column, ‘High’, we see that many values end in 0001 or 9999. This is likely to be the result of rounding errors. Let’s tidy up the column by rounding all values to two decimal places.

    Navigate to the **Transform** ribbon > **Number Column** group > **Rounding** > **Round…**. Under ‘Decimal Places’, enter the value 2 and click **OK**.



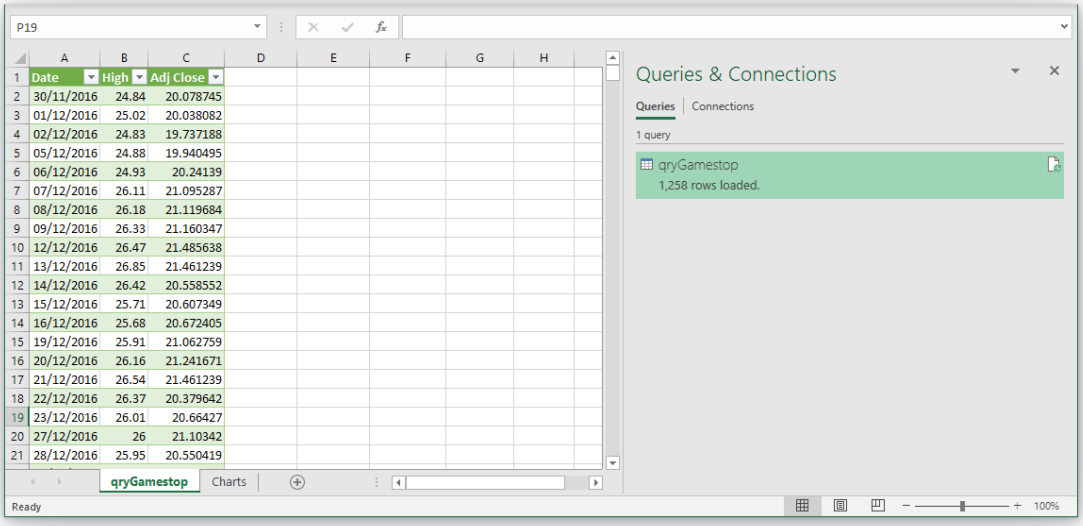

    We see that the values in the ‘High’ column have now been rounded, and a corresponding step has appeared in the **Applied Steps** list.

27. Navigate to **Ribbon** > **Home tab** > **Close group** > **Close & Load**. This will return us to Excel.

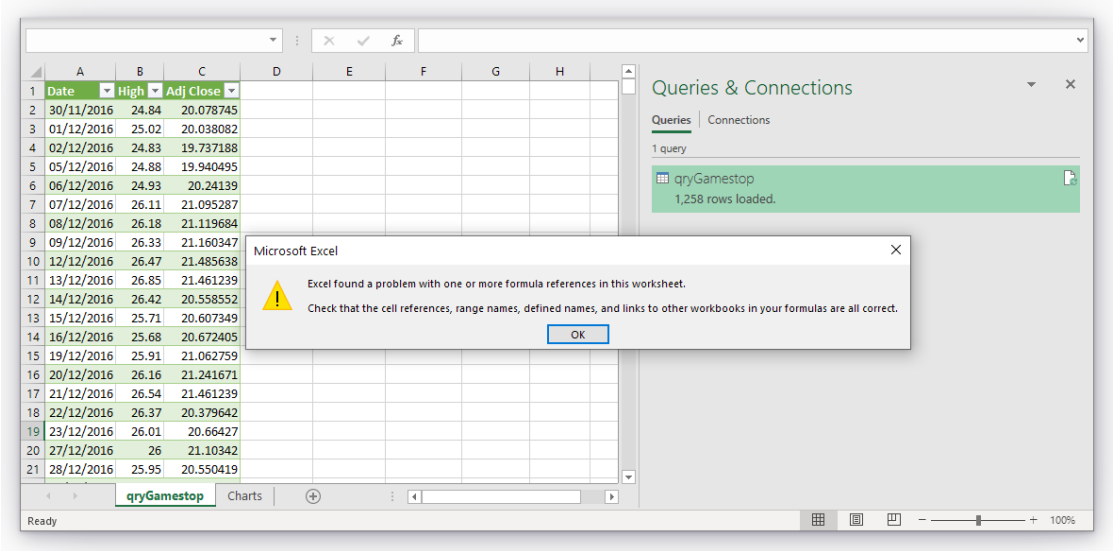


## 3.2 Back to Excel

28. Back in Excel, the **qryGamestop** table has been updated to reflect the changes we just made in Power Query. (Note that the table is linked to the Query.)




29. After a few seconds, Excel will raise the error, **'Excel found a problem with one or more formula references...'**. When this appears, click **OK**. Navigate to the **Charts** worksheet.




30. We see that the 'Volume' chart is now broken. This is expected since we removed the 'Volume' column from the **qryGamestop** Query. Consequently, the 'Volume' column of the **qryGamestop** table has disappeared.

31. Delete the 'Volume' chart.

⚠ When we alter a Query, we must take care to consider the effects this might have on any associated Excel workbooks. N.B. if we were now to edit the **qryGamestop** Query to restore the 'Volume' column, this would **not** fix the 'Volume' chart: the broken data link is broken permanently. The only recourse would be to undo (**Ctrl + Z**) the changes.

32. To refresh a Query, we have several options:

    - Right-click the **qryGamestop** table > **Refresh**.
    - Open the Queries & Connections panel (**Ribbon** > **Data** tab > **Queries & Connections** group > **Queries & Connections**), right-click the name of the Query > **Refresh**.
    - **Ribbon** > **Data** tab > **Queries & Connections** group > **Refresh All**. Note: this will refresh all Queries in the workbook and all Pivot Tables.

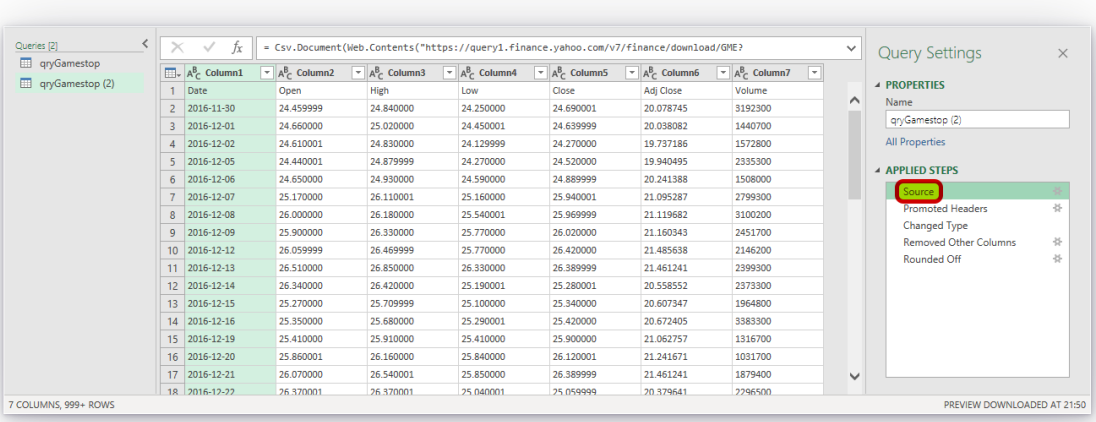
    Refresh the **qryGamestop** Query. You may see some of the values in the **qryGamestop** table change as a result.

## 3.3 Edit the source URL

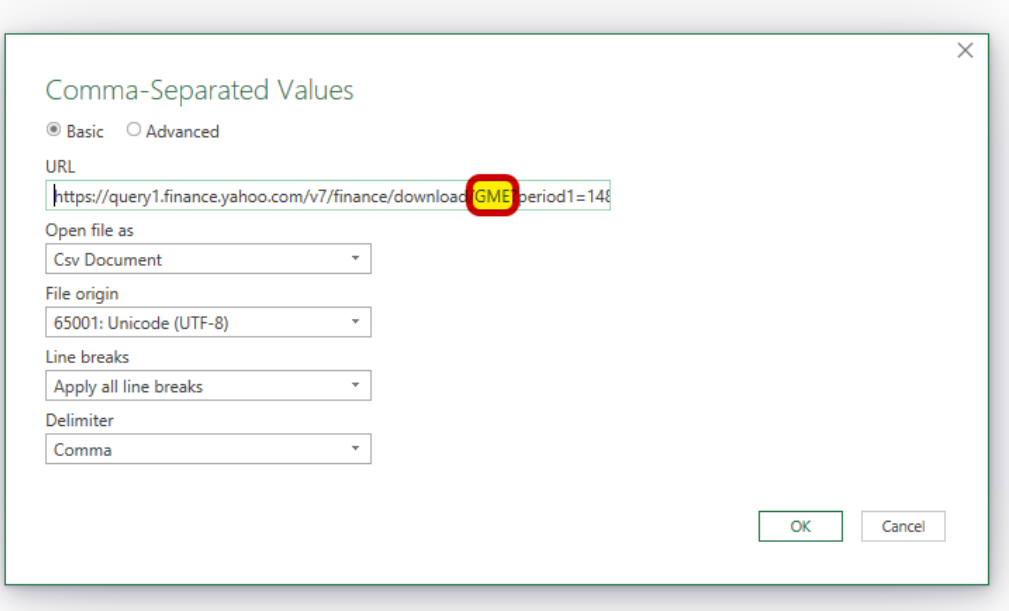
We will duplicate our Gamestop Query and edit the copy so that it retrieves the price history of a different stock. For Yahoo Finance, this can be achieved by making a small change to the source URL that we copied at the start of this exercise.

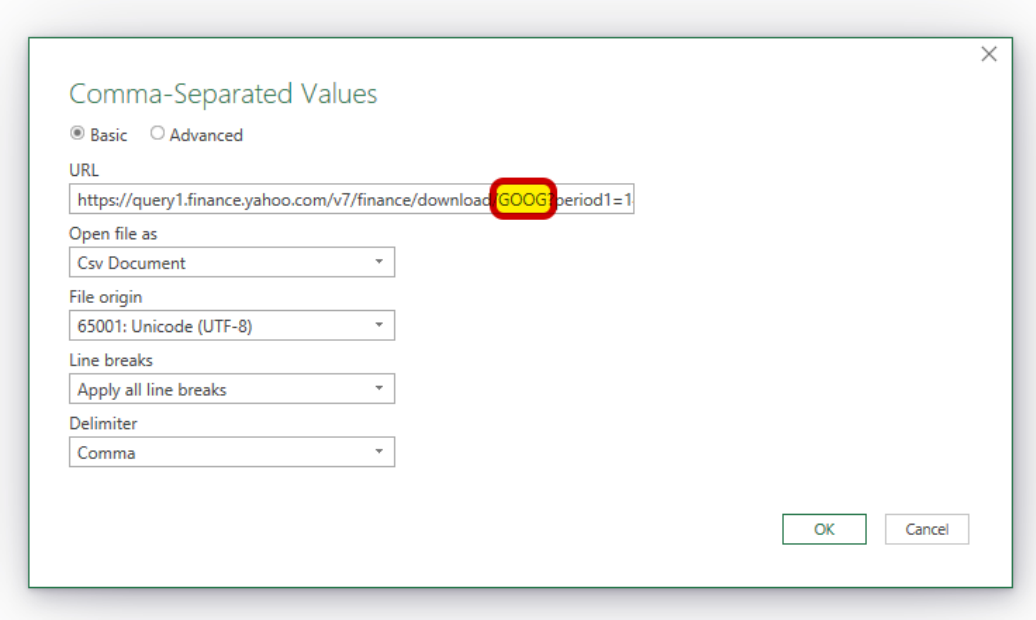
33. Create a duplicate of the **qryGamestop** Query by right-clicking the Query in the **Queries & Connections** panel and selecting **Duplicate**. The Power Query Editor window will open as below. Note that a new Query, 'qryGamestop (2)' has appeared in the **Queries** panel on the left-hand side of the Power Query Editor window. Ensure that 'qryGamestop (2)' is highlighted.

34. In the **Applied Steps** list for Query **qryGamestop (2)**, double-click the first step, ‘Source’.




35. A dialog called ‘Comma-Separated Values’ opens, as below. Edit the URL field by replacing your original stock symbol (here, GME) with a different stock symbol, for example MSFT (Microsoft), AMZN (Amazon) or GOOG (Google).



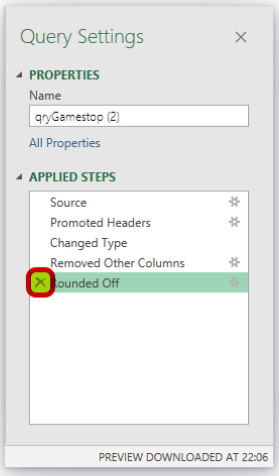



36. Allow a few moments for the Query to update, then click **OK**.

37. Now we will make a quick visual check to verify that our modified Query works as expected.

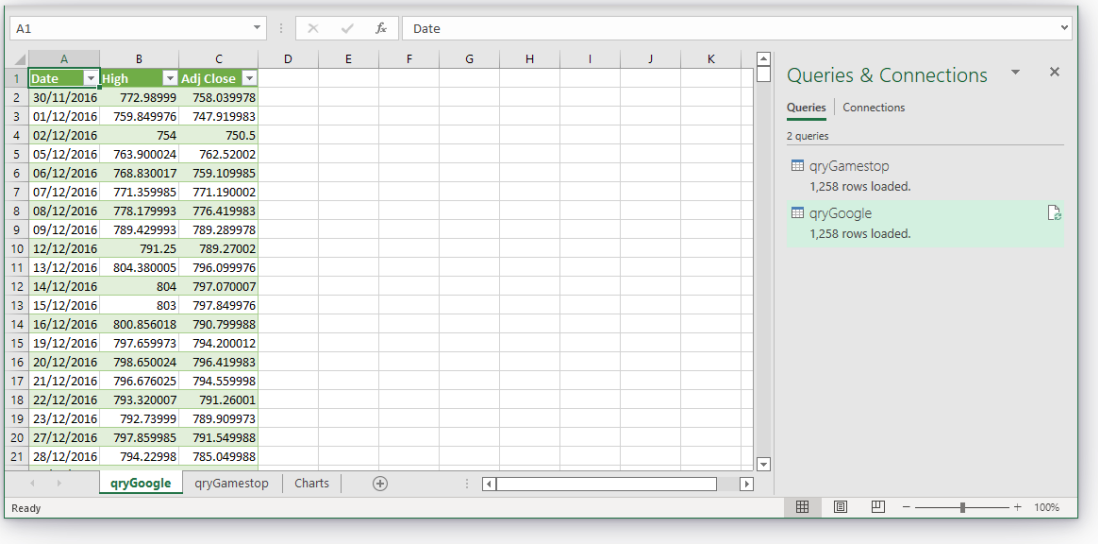
    In the **Applied Steps** list, click on each of the following steps in turn: 'Promoted Headers', 'Changed Type', and 'Removed Other Columns'. Examine the data after clicking each step.

    If everything looks fine, proceed to the 'Rounded Off' step. Recall that this step rounds the 'High' column to two decimal places. If this transformation does not seem appropriate for the current Query, delete the step by clicking the ✗.




38. In the Queries tab, right-click the Query named **qryGamestop (2)** > **Rename**. Rename the Query as appropriate. Click **Close & Load**.

39. We now have two independent Queries, each linked to a different share price.




| Date | High | Adj Close |
|---|---|---|
| 30/11/2016 | 772.98999 | 758.039978 |
| 01/12/2016 | 759.849976 | 747.919983 |
| 02/12/2016 | 754 | 750.5 |
| 05/12/2016 | 763.900024 | 762.52002 |
| 06/12/2016 | 768.830017 | 759.109985 |
| 07/12/2016 | 771.359985 | 771.190002 |
| 08/12/2016 | 778.179993 | 776.419983 |
| 09/12/2016 | 789.429993 | 789.289978 |
| 12/12/2016 | 791.25 | 789.27002 |
| 13/12/2016 | 804.380005 | 796.099976 |
| 14/12/2016 | 804 | 797.070007 |
| 15/12/2016 | 803 | 797.849976 |
| 16/12/2016 | 800.856018 | 790.799988 |
| 19/12/2016 | 797.659973 | 794.200012 |
| 20/12/2016 | 798.650024 | 796.419983 |
| 21/12/2016 | 796.676025 | 794.559998 |
| 22/12/2016 | 793.320007 | 791.26001 |
| 23/12/2016 | 792.73999 | 789.909973 |
| 27/12/2016 | 797.859985 | 791.549988 |
| 28/12/2016 | 794.22998 | 785.049988 |

# 4 Summary

In this activity, we have:

- created a Query in Microsoft Excel via the **Get Data** > **From Web** wizard.
- learnt how to rename, refresh, and duplicate Queries.
- gained a basic familiarity with the Power Query Editor window.
- learnt a little about some of the (many) features of Power Query.
- added, edited, and deleted transformation steps within a Query.

qa.com