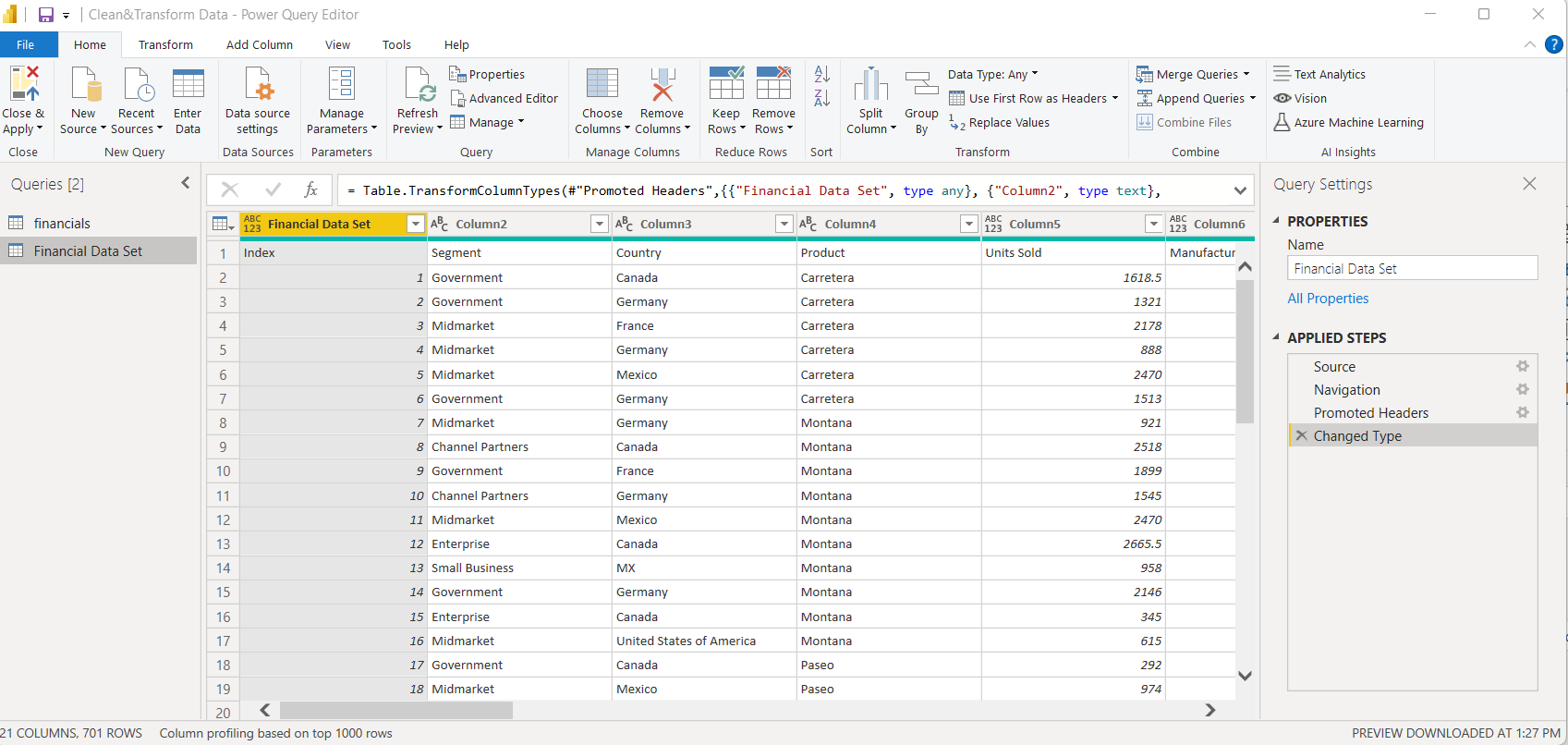
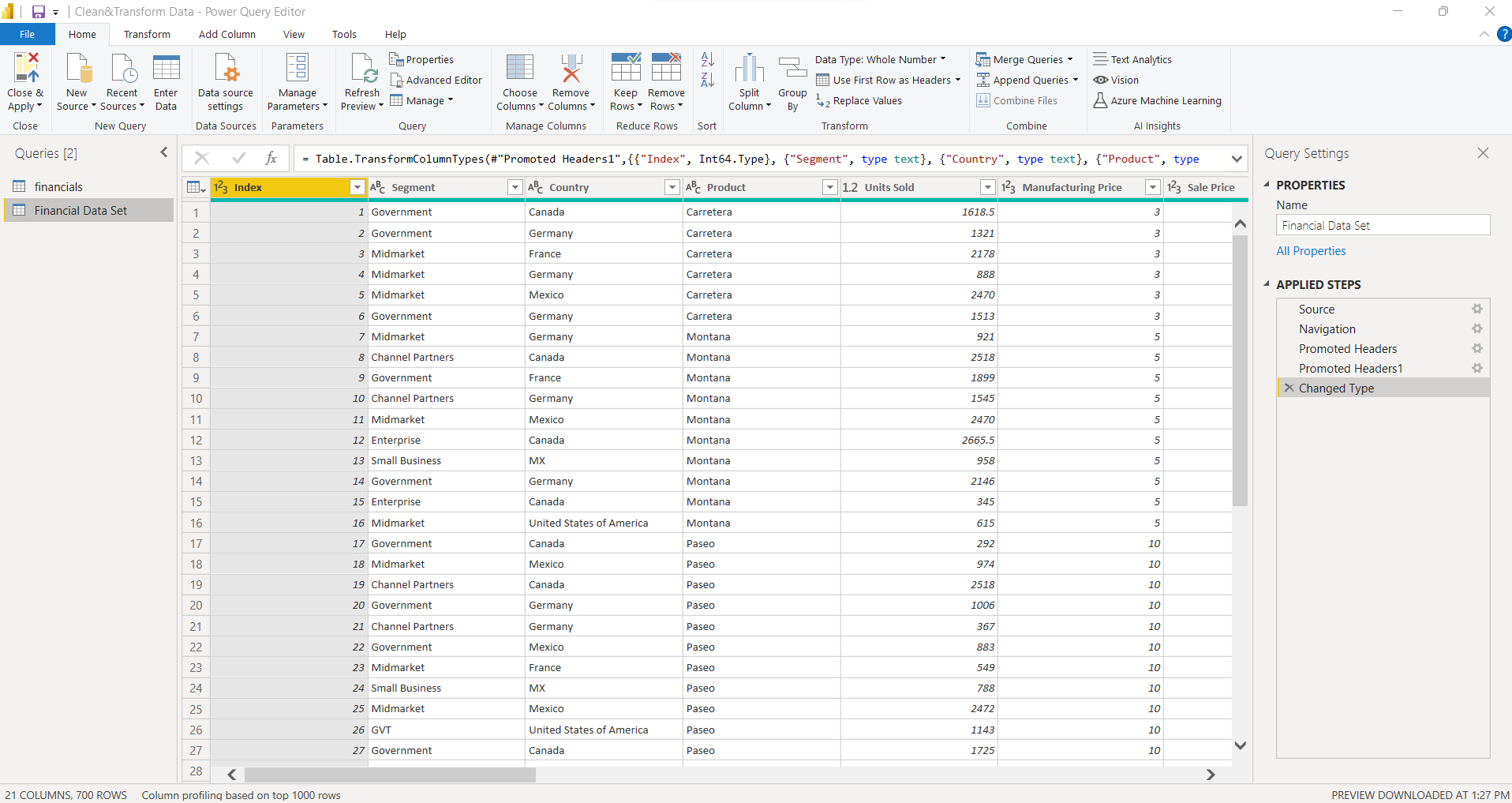
1. Describe the way(s) that the data set does not comply with "tidy data" principles. How would you prepare the data so that they were tidy for analysis? Apply the fix you describe to the data set.

The column headers are on the second row instead of the first row. Select Use first row as headers in the Transform group to insert the column headers on the first row.

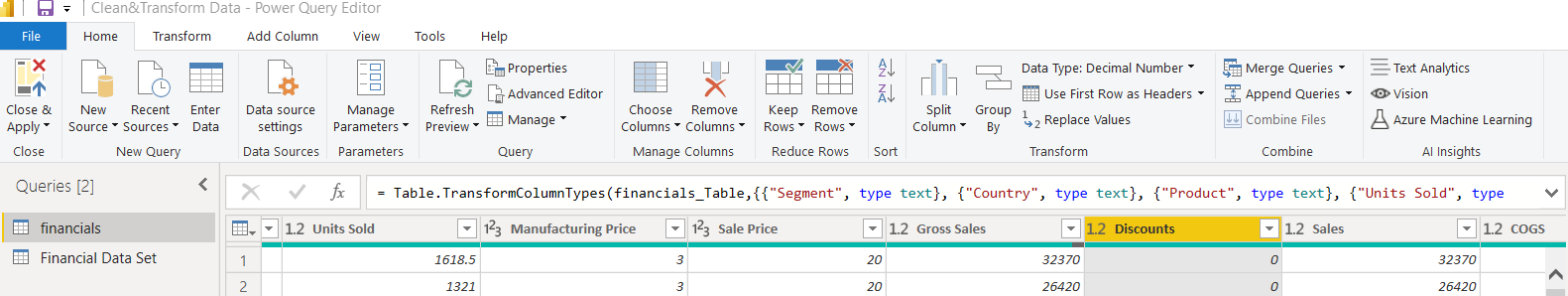


Fixed data set:

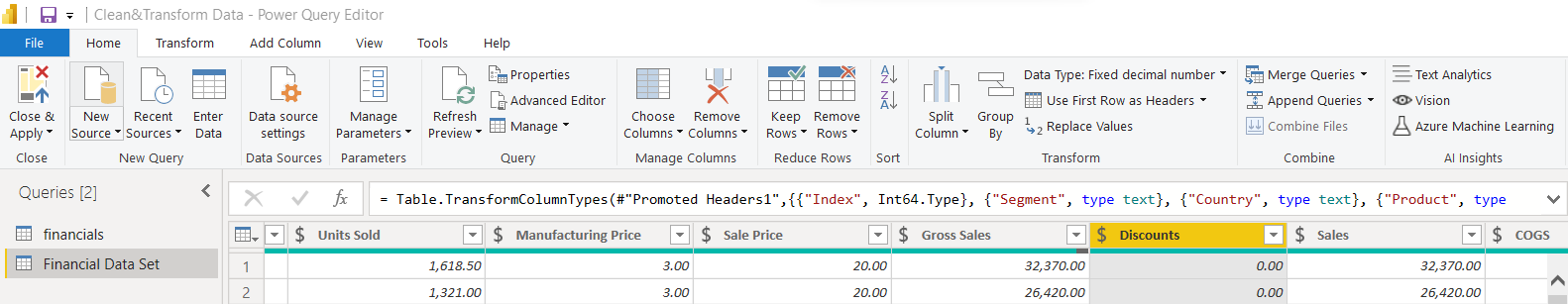


1. Describe other problems with the data and the associated fixes for them. Apply these fixes to ensure the data are clean.

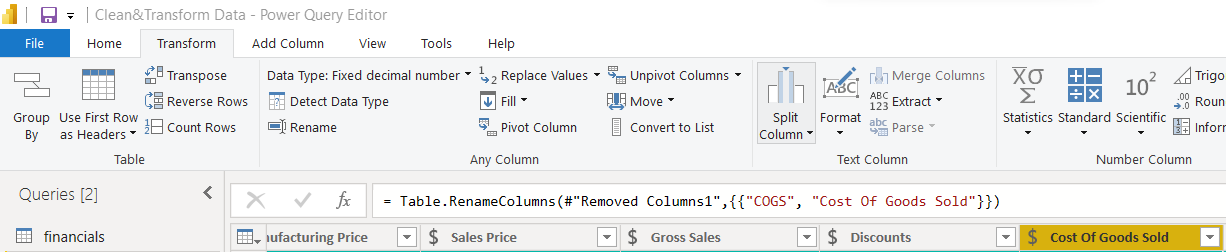
Incorrect data types for columns units sold, manufacturing price, sale price, gross sales, discounts, sales, cogs, and profit.

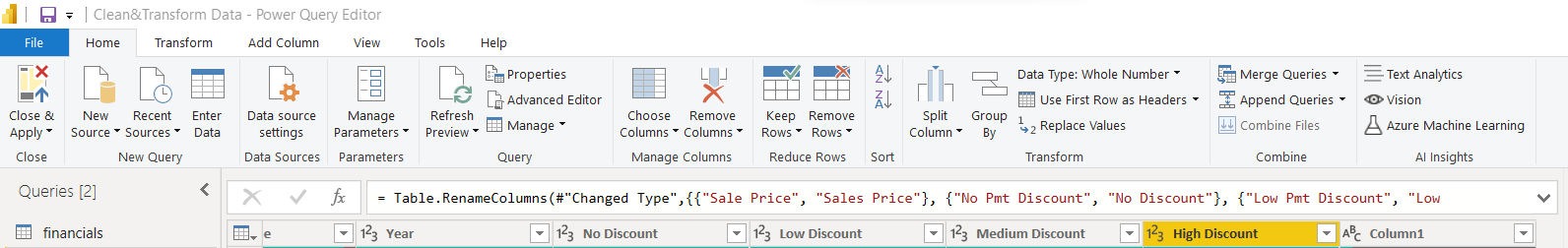


Change data types from decimal numbers and whole numbers to fixed decimal numbers.

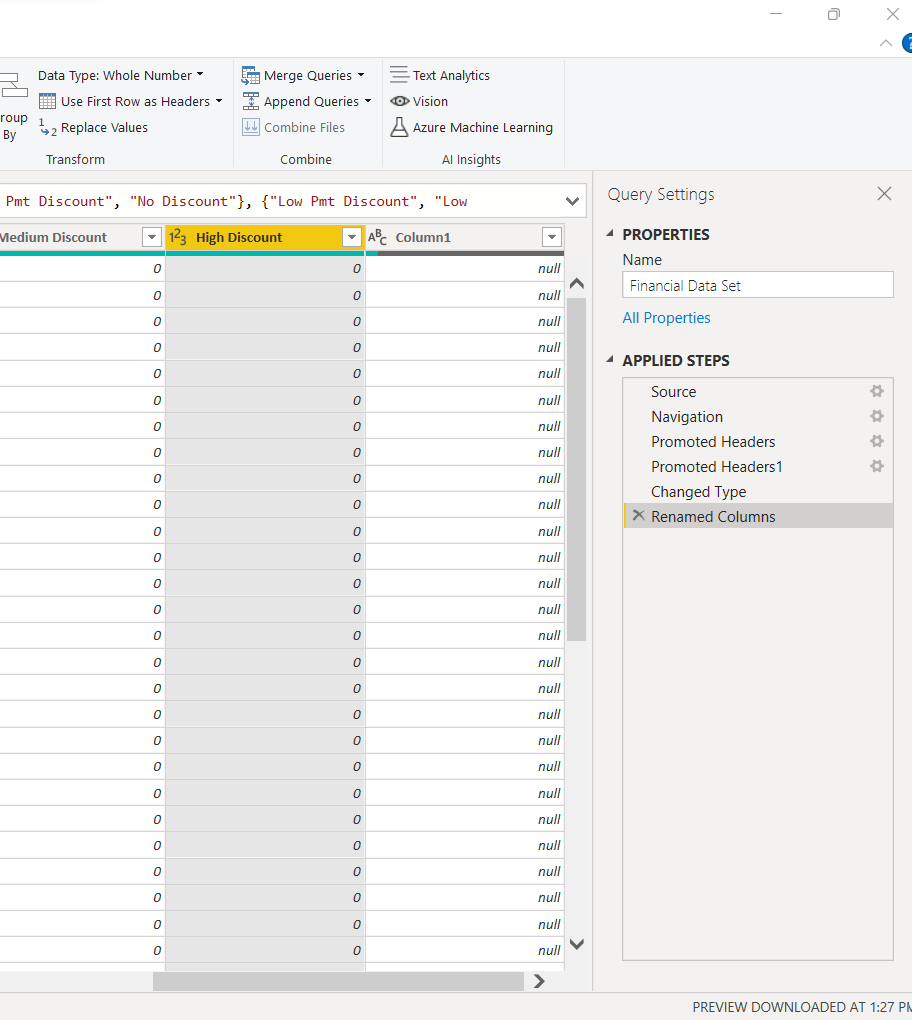


Rename column names Sale price to Sales price, COGS to Cost Of Goods Sold, Discount columns to No discount, low discount, medium discount, and high discount.

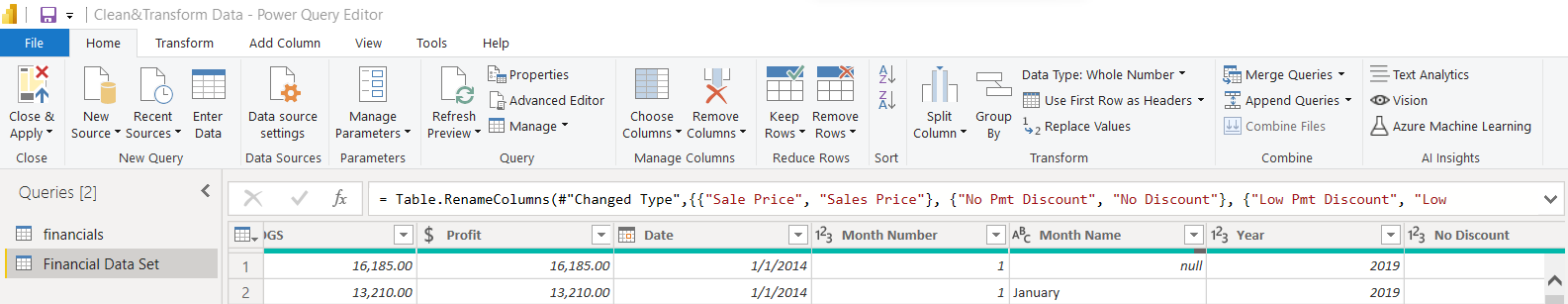


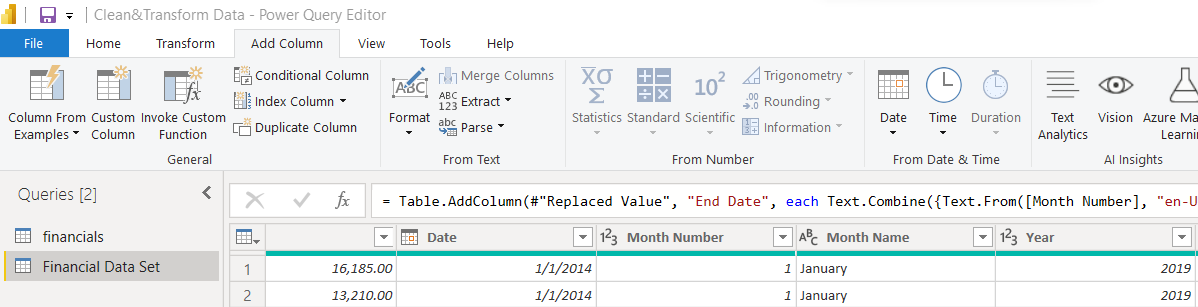


Remove unnecessary columns, Column 1.

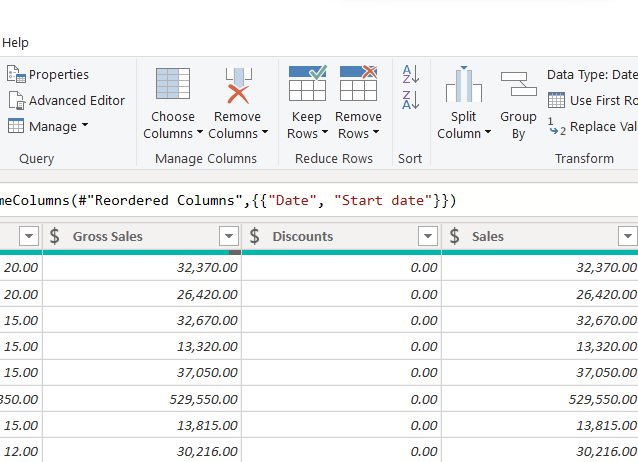


Replace null value with month name January under the month name column.





Delete one sales column because they both contain the same data (i.e., Sales and gross sales).



Combine month number, month name, and year columns to create a date column using the Column from examples option in the General group and rename it End date.

Rename the date column to Start date.

Replace wrong/invalid dates with valid ones.

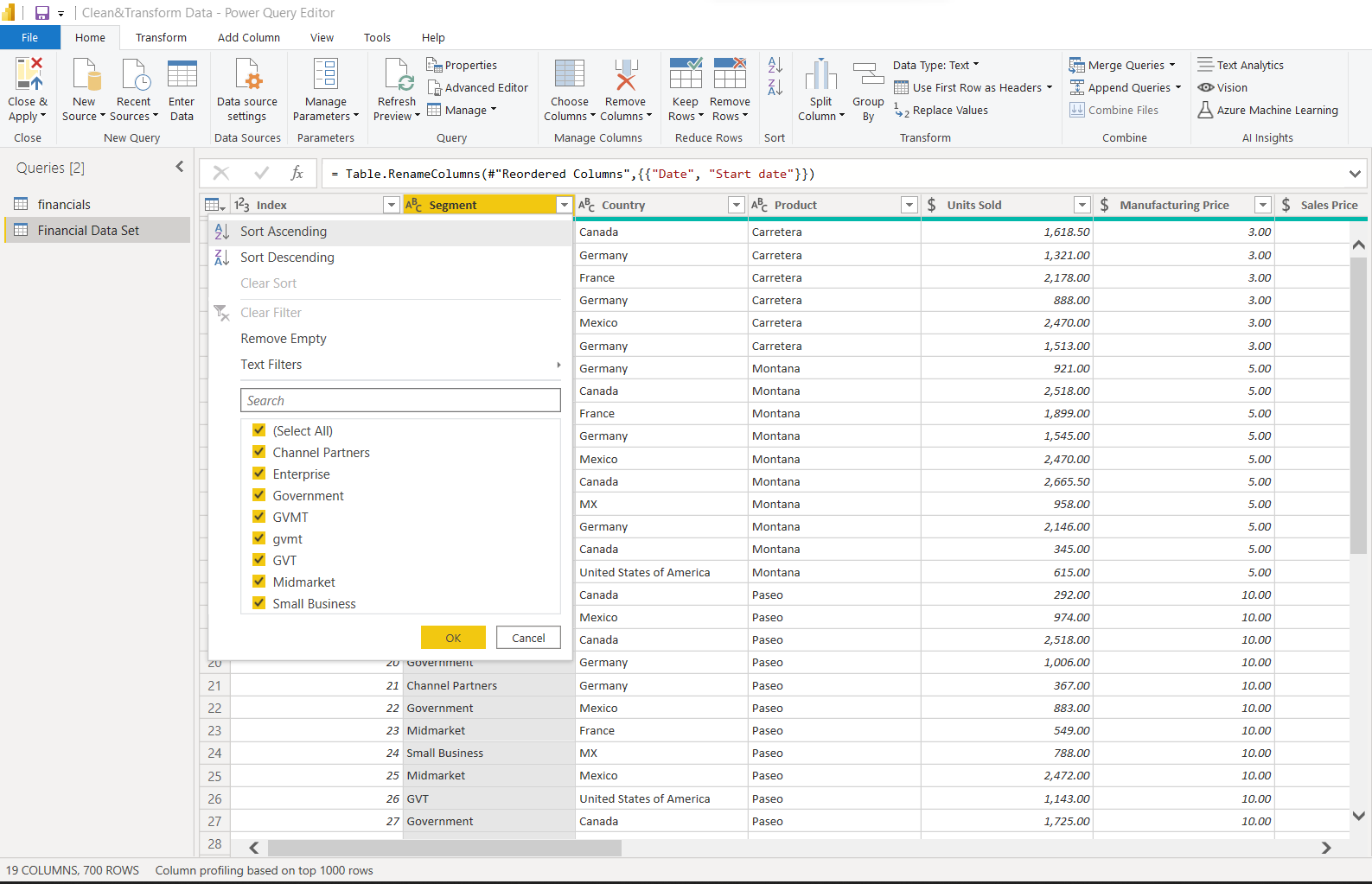
Delete month number, month name, and year columns.

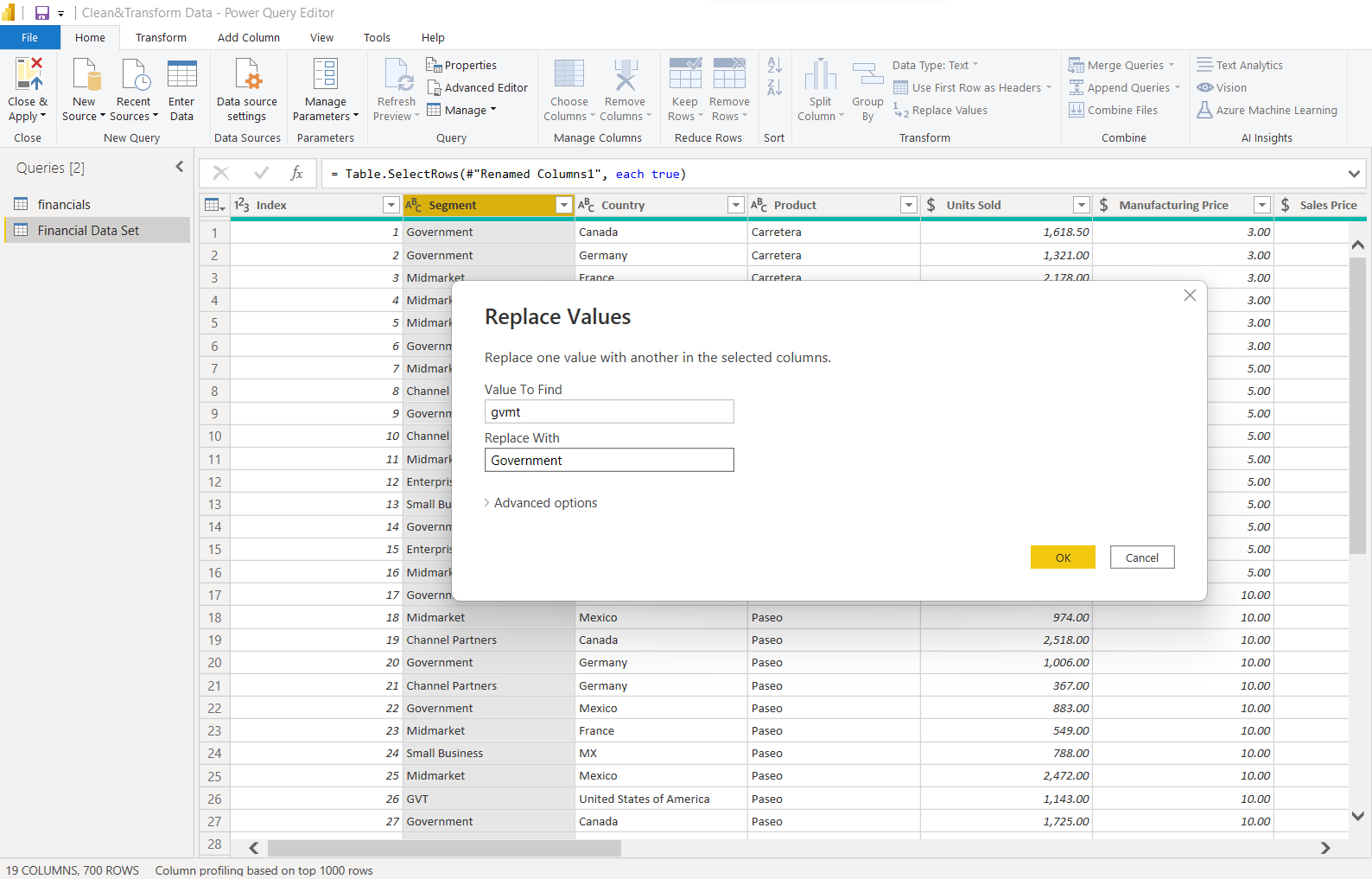
Change the End date data type to date.

Graphical user interface, application

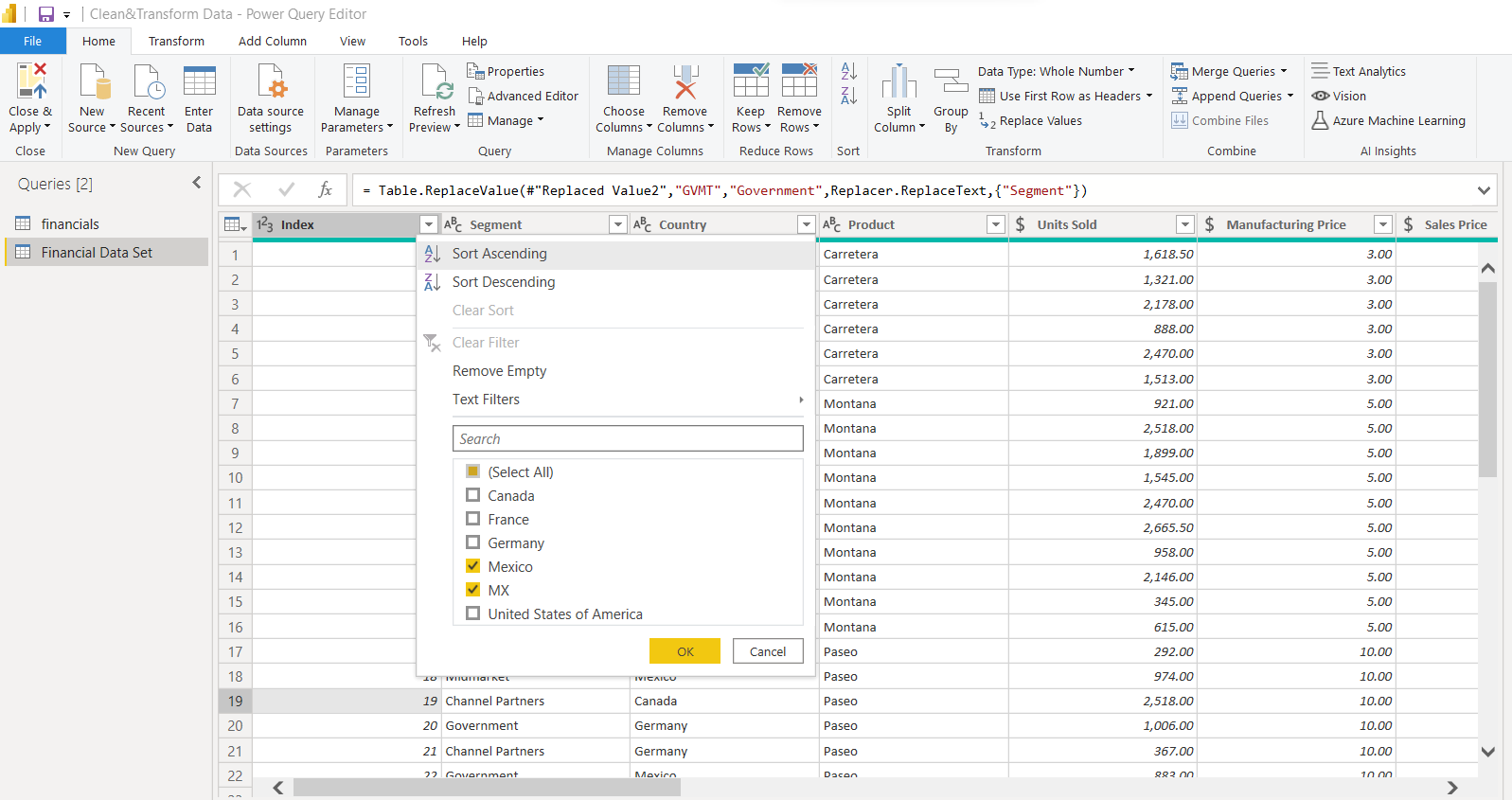
Description automatically generated

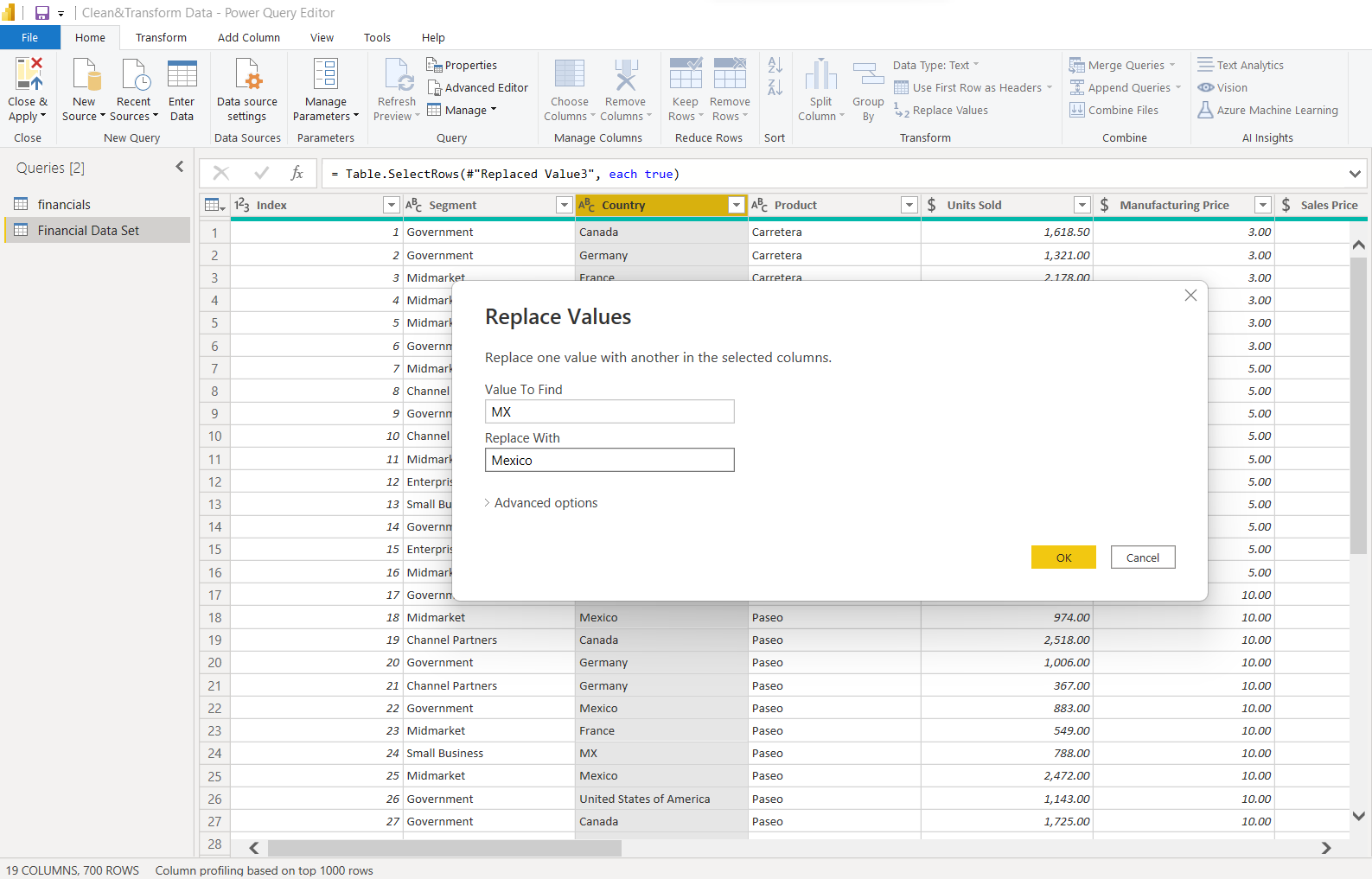
Replace the abbreviation of segment gvmt, GVMT, and GVT with government under the segment column.



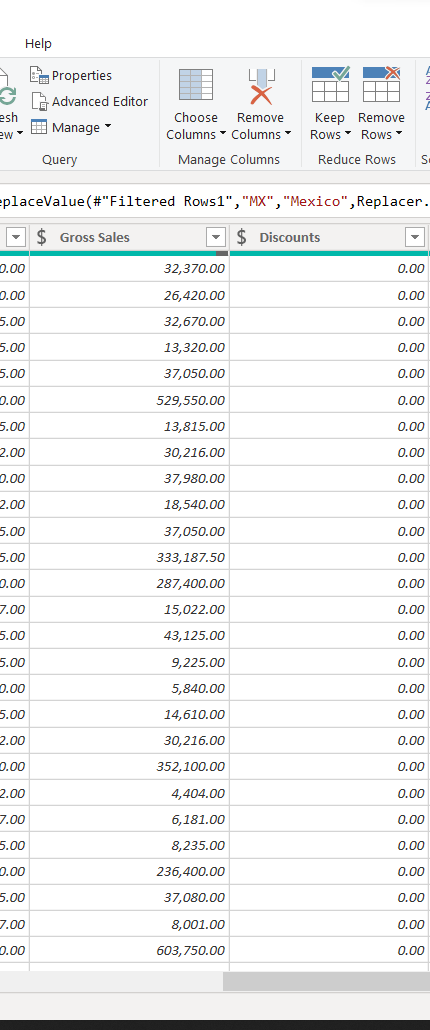


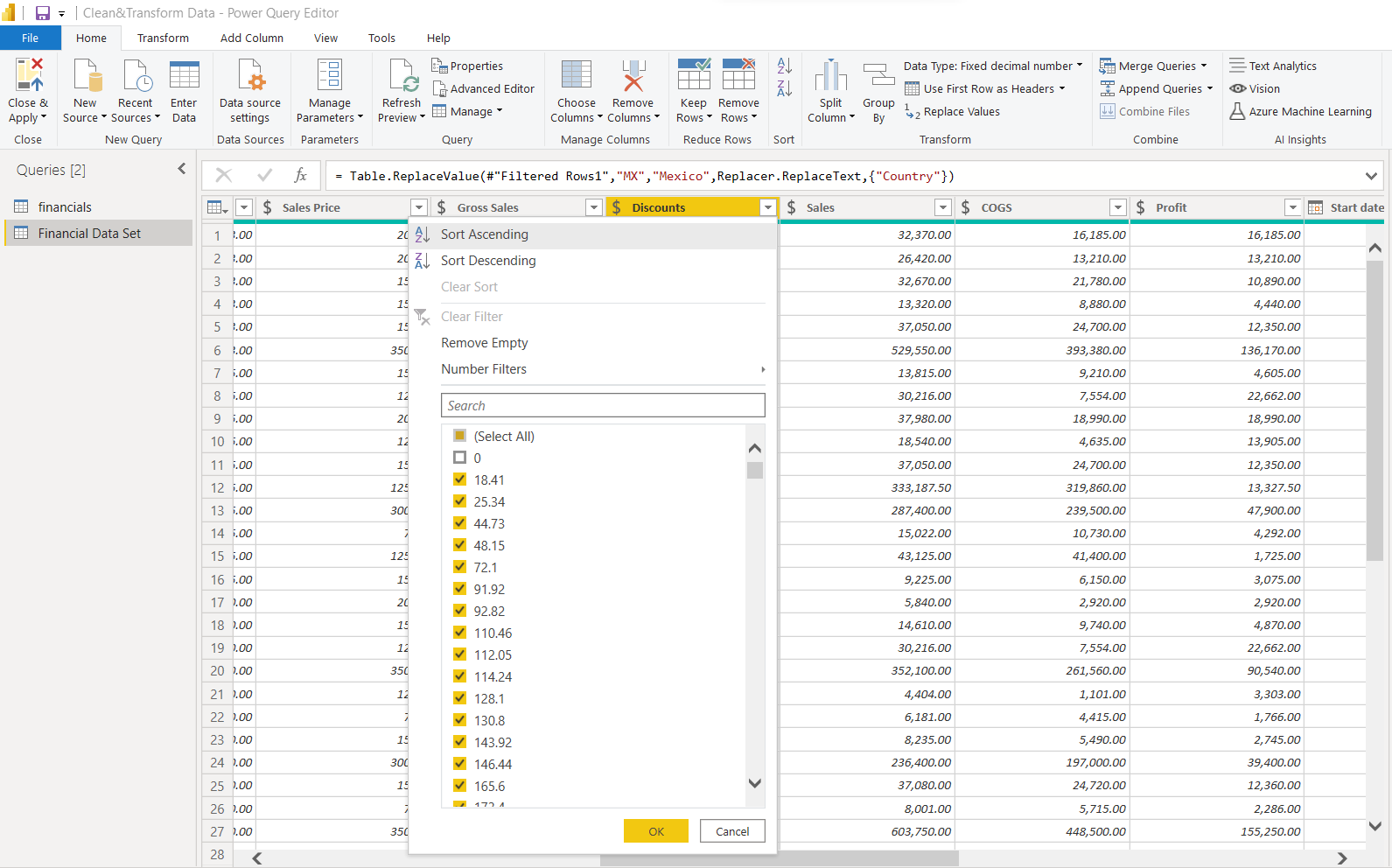
Replace MX with Mexico under the country column.



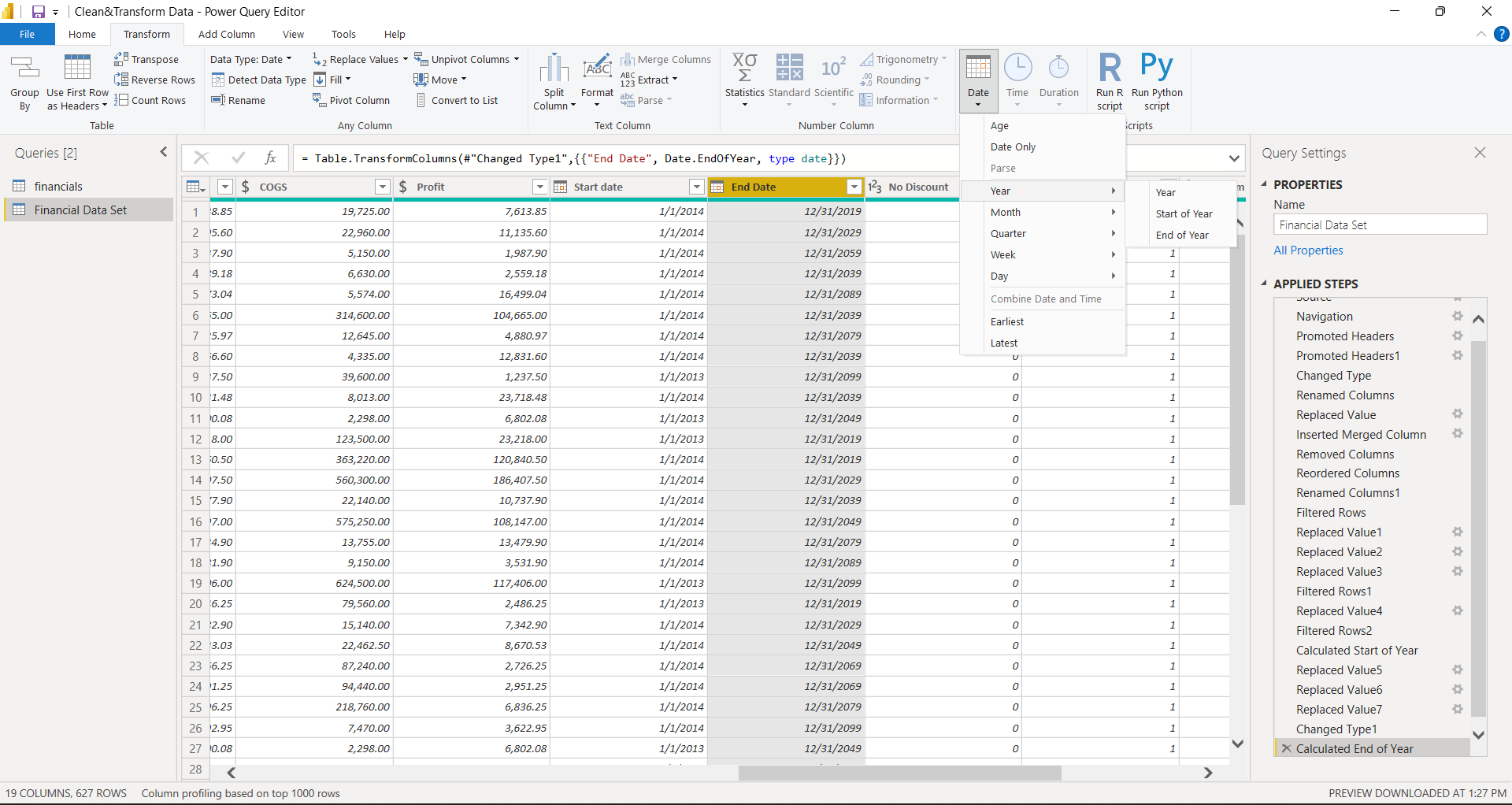


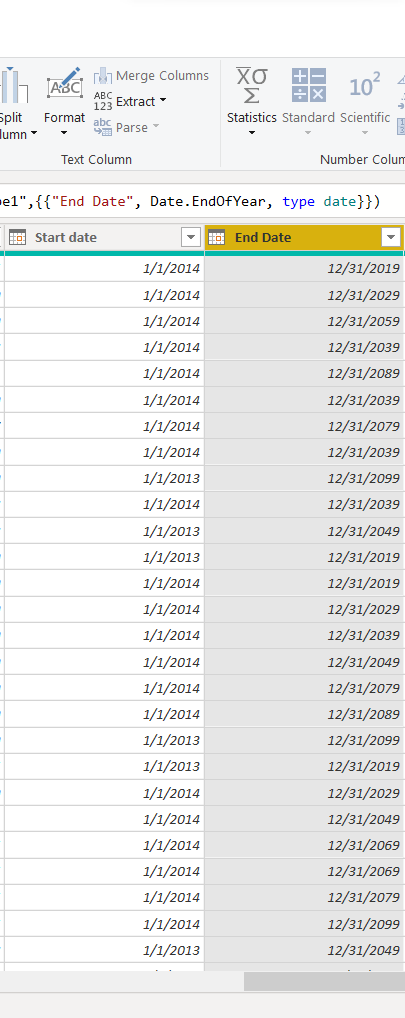
Remove the value 0 under the discounts column.





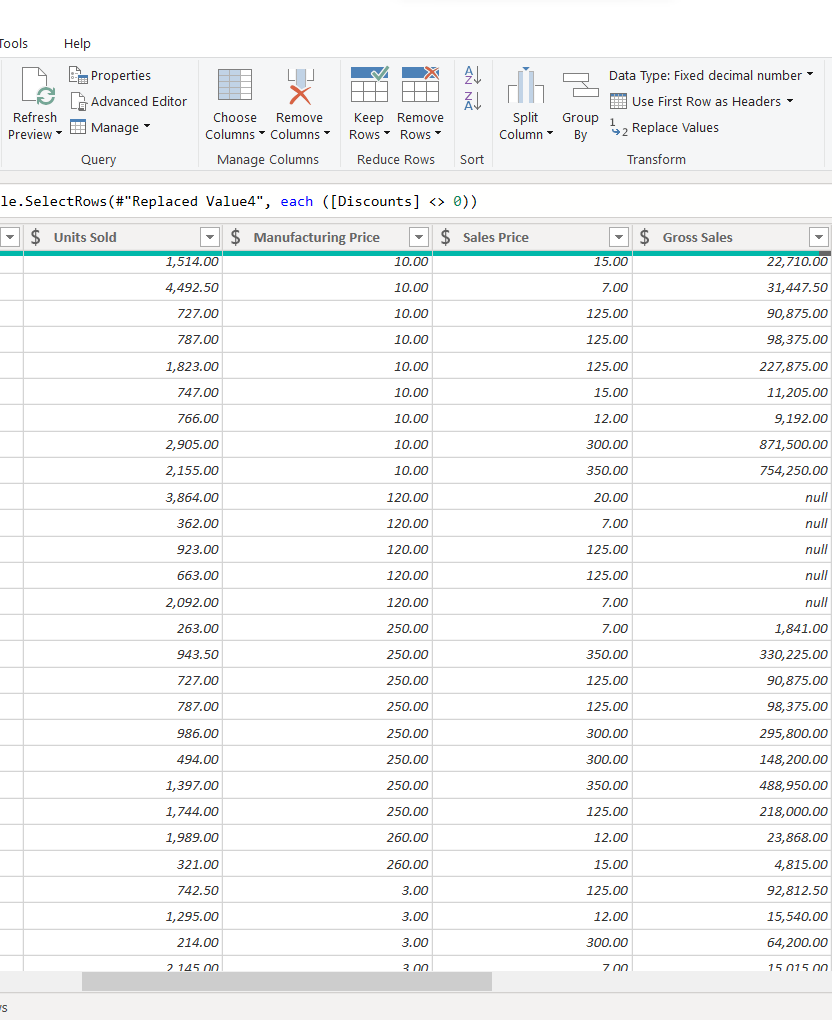
1. In an interview with the head of the Channel Partners segment, you learn that they code their sales data on a fiscal-year basis rather than a calendar year, as the other segments do. Prepare the data such that all segments, including Channel Partners, report their sales data on a calendar-year basis.

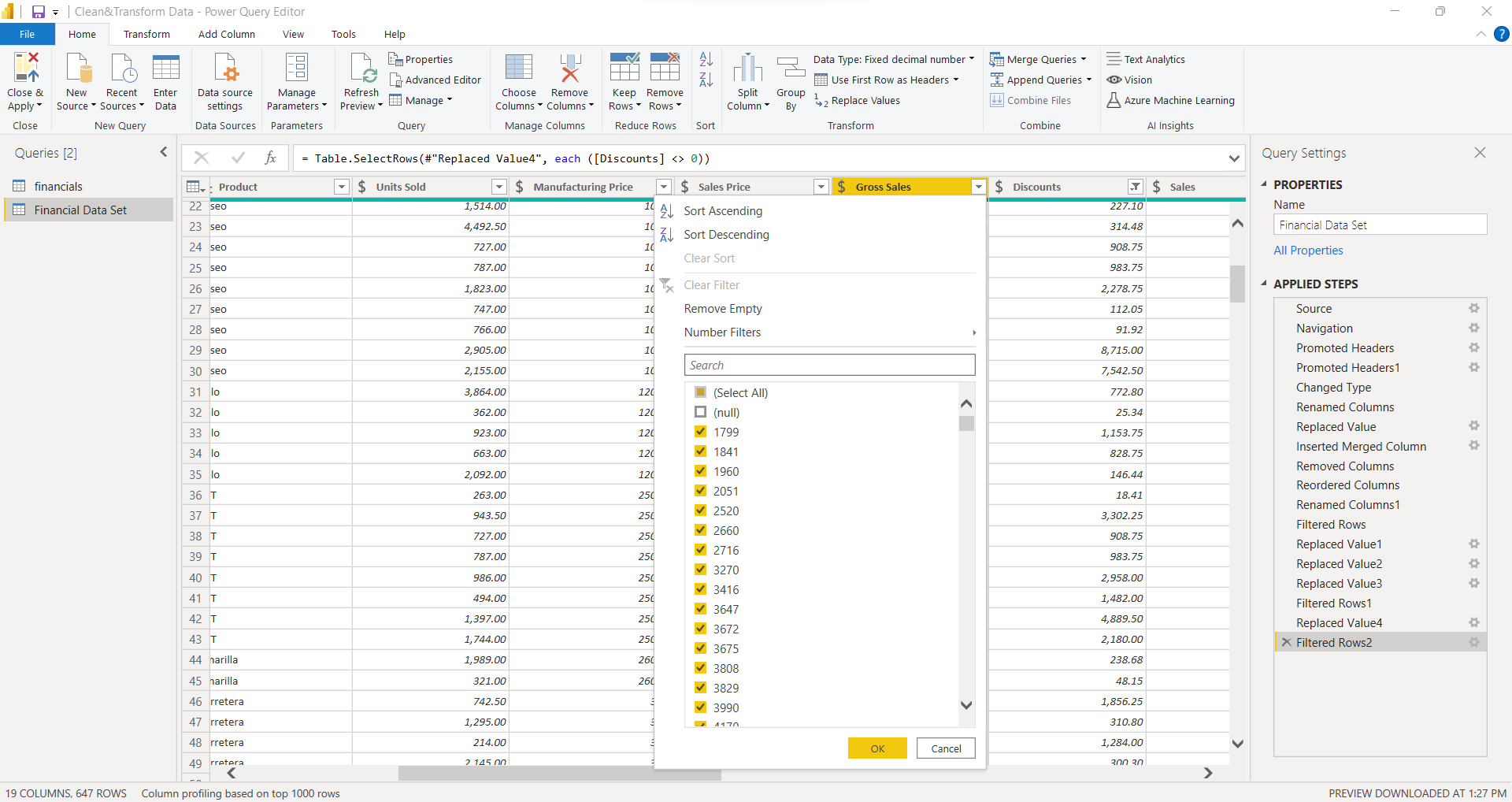




1. Are there any unusual values in the data set for the variables Units Sold, Sales Price, Manufacturing Price, or Gross Sales? If so, what is the unusual value. Explain how to deal with any unusual values for purposes of preparing the data for analysis.

The Gross sales column has null values. They can be removed by using the filter option next to the column header.





1. Calculate the correlation between (a) the units sold and profit and (b) sales price and profit. Which is more strongly correlated to profit, units sold or sales price? Provide an explanation with support.

Graphical user interface, application, table

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, Word

Description automatically generated

Sales price is more strongly correlated with profit because the data set shows that the higher the sales price, the higher the profit.