

Display the **Sales Orders** data in the Query Editor.

- a. Select **Home**→**Transform data**→**Transform data**.
- b.

The screenshot shows the Microsoft Power Query ribbon with the 'Home' tab selected. In the 'Transform' section of the ribbon, the 'Transform data' button is highlighted with a yellow box. Below the ribbon, there is a preview of three rows of data from a 'Sales Orders' table, showing columns for Order ID, Order Date, and Ship Date.

- c. In the **Power Query Editor** window, in the **Queries** pane, select the **Sales Orders** table.

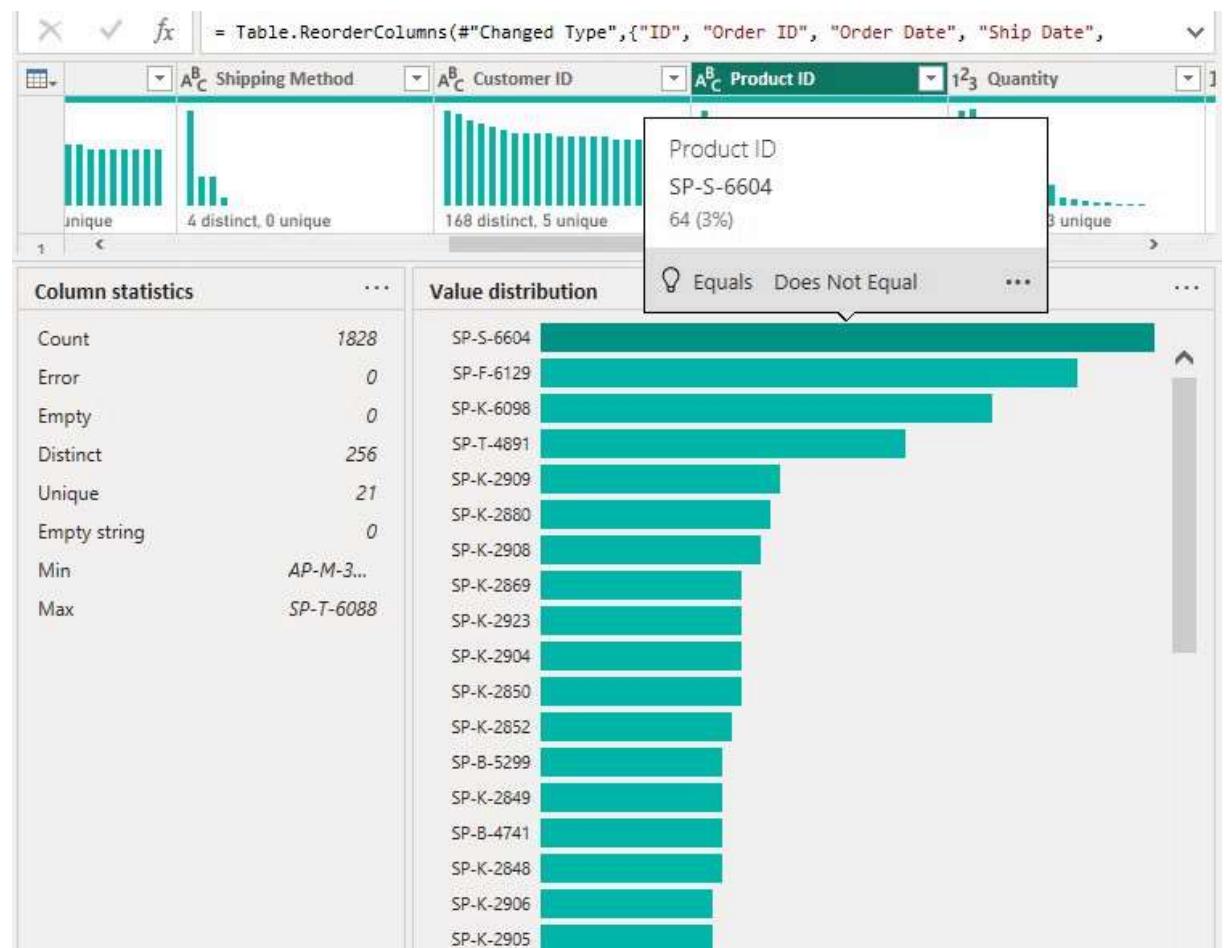
The screenshot shows the Power Query Editor window. The 'Queries' pane on the left has 'Sales Orders' selected, highlighted with a yellow box. The main area shows a preview of the 'Sales Orders' table with columns 'ID' and 'Order ID'. The 'View' tab is selected at the top, and the 'Column profile' checkbox is checked, also highlighted with a yellow box.

Open the **Column Profiles** pane, and examine the profiling information provided by Power BI.

- a. Select **View** and check the **Column profile** check box.

The screenshot shows the Power Query Editor window with the 'View' tab selected at the top. The 'Column profile' checkbox in the 'View' settings is checked, highlighted with a yellow box.

Evaluate Product ID Column



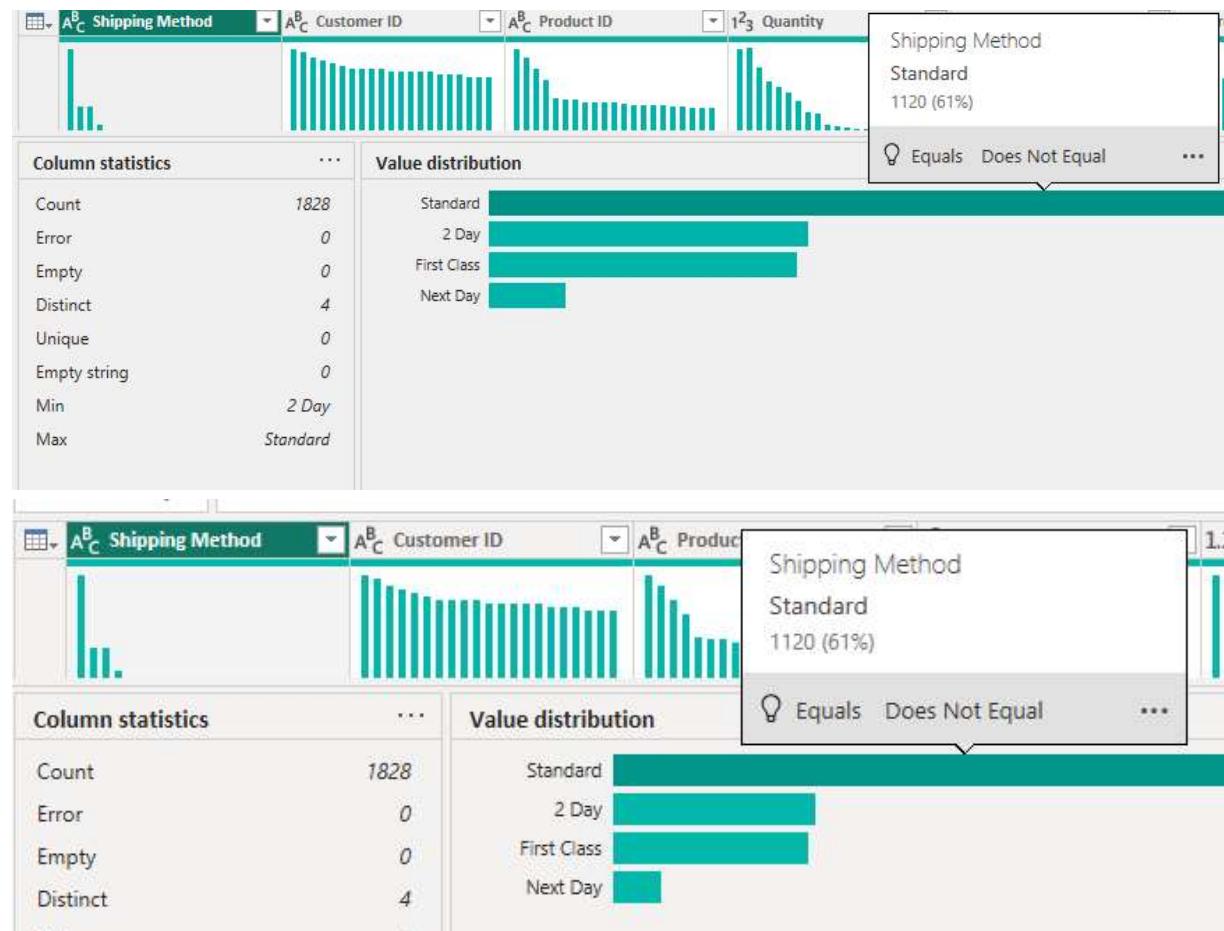
Statistic	Value
Count	1,828
Distinct	256
Unique	21

Most Frequent ID SP-S-6604 (64 times, or 3% of transactions)

Data Quality 0 Errors, 0 Empty

It has low cardinality (256 distinct IDs) relative to the total number of records, making it an excellent dimension for grouping and filtering your sales data. The ID SP-S-6604 is the most frequently sold product and should be a focus for inventory and sales performance analysis.

Evaluate Shipping Method



Observation:

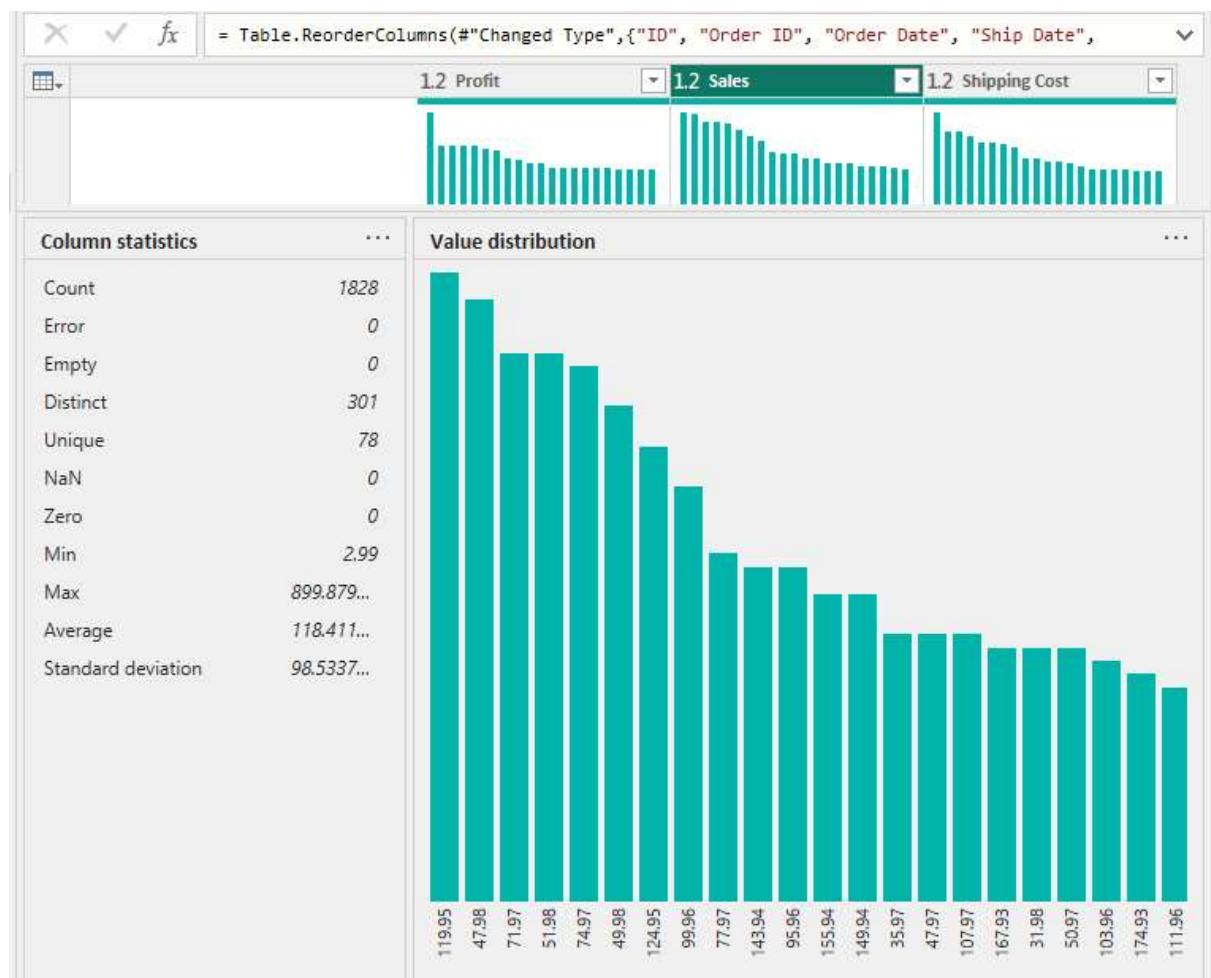
Standard: This is the **most frequent** shipping method, dominating the column's values.

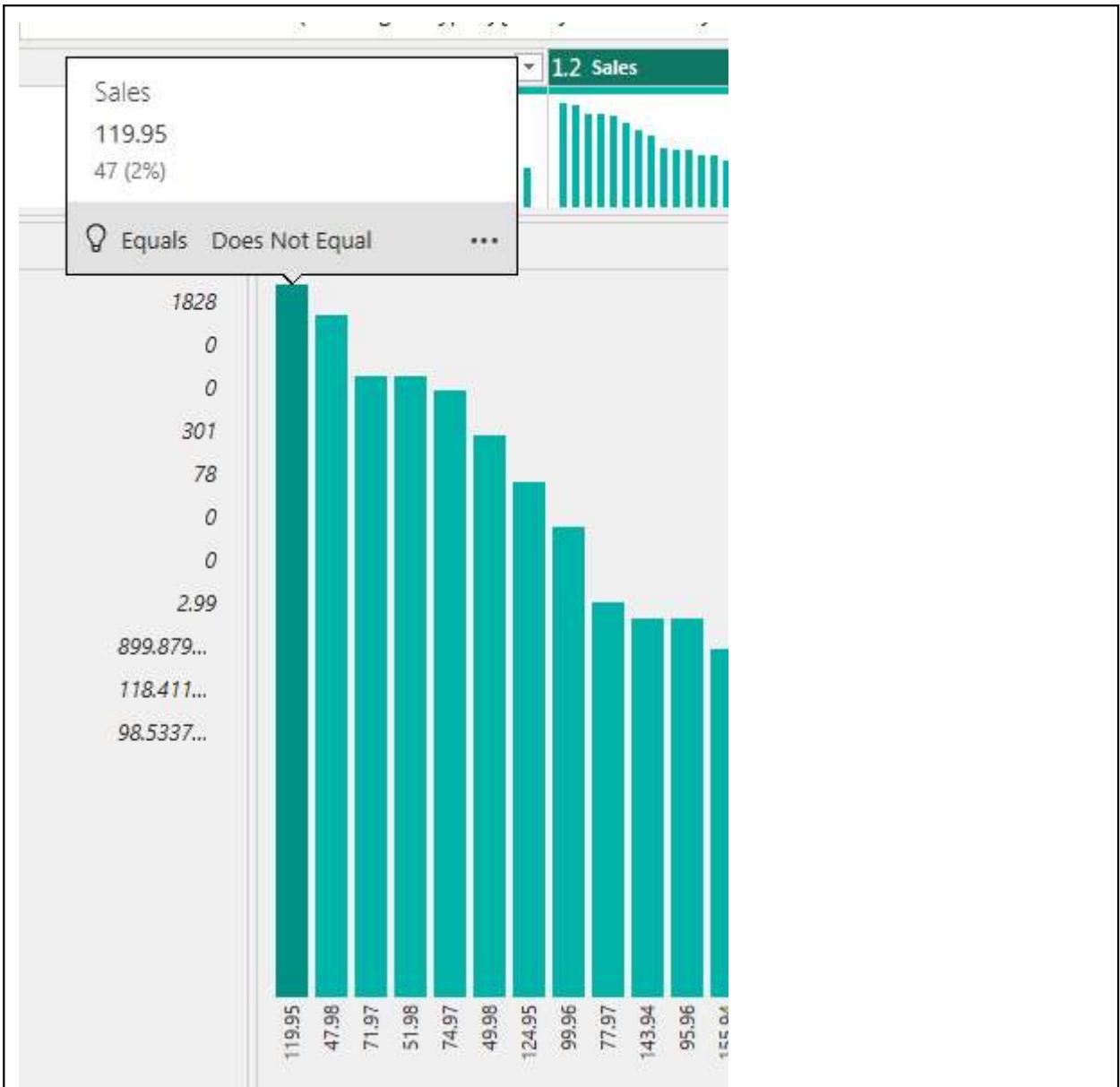
2 Day and First Class: These two categories are similarly frequent, representing a significant but smaller portion of the transactions.

Next Day: This is the **least frequent** category, indicating it is rarely used or possibly an expensive/premium option.

Out of a total of 1828 orders, 1120 (61%) used the Standard shipping method

Evaluate Sales Column





Extracting the key statistics:

Min value sale is \$2.99

Max value sale is \$899.88 (Outlier)

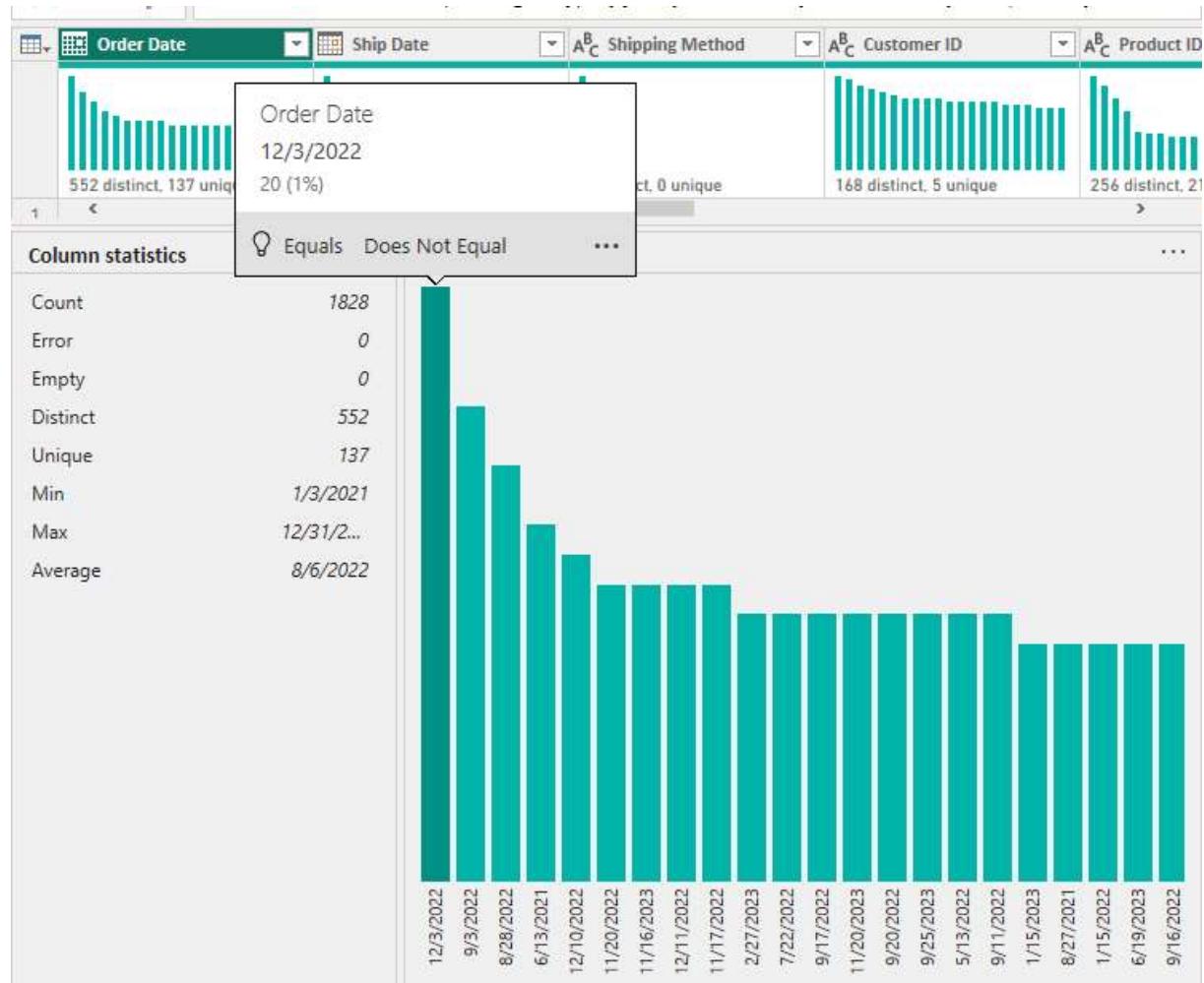
Most Frequent sale amount is \$119.95

Average Sale \$118.41

Standard Deviation \$98.53

Interpretation: I have a large Standard Deviation (SD) which indicates the presence of outliers (or at least a very highly skewed distribution) in your

data.



Count 1,828

Min Date 1/3/2021

Max Date 12/31/2023...

Distinct Dates 552

Most Frequent Date 12/3/2022 (20 transactions, or 1% of total)

Average Date 8/6/2022

Data Quality 0 Errors, 0 Empty

Interpretation

The Order Date column has perfect data quality (0 Errors, 0 Empty) and spans nearly three years, from early 2021 to late 2023.

With 552 distinct dates, this column is essential for time-series analysis to track trends over months and years. The Average Date of 8/6/2022 and the value distribution show that the data is concentrated more recently, with the latest dates (like 12/3/2022) having the highest transaction counts, suggesting a growth or recent surge in activity.

In summary

Interpretation:

Time series columns (like Order Date) are excluded from the *low-cardinality dimension* group because they are handled differently in analysis. The Sales column is considered high cardinality only in comparison to the other columns in my model.

1. High Cardinality: With 301 distinct values out of 1,828 records, Sales has a much higher cardinality than your other columns, such as:
 - Shipping Method (4 distinct values)
 - Product ID (256 distinct values)
2. High Variability: The Standard Deviation (\$\$\$98.53) is nearly equal to the Average (\$\$\$118.41). This indicates extreme spread in the data, with the Max value of \$\$899.88 acting as a significant outlier.
3. Data Quality: The column is perfectly clean (0 Errors, 0 Empty) and ready for calculations.

Conclusion for Dashboard Use

Because the Sales column is a high-cardinality measure with high variability, it should be used for calculations (SUM, AVG, COUNT) and should always be viewed segmented by your low-cardinality dimension columns, such as shipping method or product ID

