Changelog Documentation

Project Details	
Client	Retail Company
Project	Retail Company Sales Performance Analysis
Start Data	2023-10-26
End Date	2023-11-25
Email	Alaamhassan2001@gmail.com
GitHub	https://github.com/alaamhassan
LinkedIn	www.linkedin.com/in/alaamhassan
Portfolio	

Table of Contents

Metadata

Field	Type	Description
Consumer ID	Text	Unique identifier for each customer.
Order ID	Text	Unique identifier for each order, where the first part is the Consumer ID and the second is a unique alphabetic char.
Month	Text	The month when the order was placed.
Year	Text	The year when the order was placed.
Total order value	Number	The total value of the order before any discounts are applied.
Discount	Number	The amount of discount applied to the order.
Line Value (net discount)	Number	The value of the order after the discount has been applied.
Line Category	Text	The category of the product in the order.
Line SKU	Text	Unique identifier for each product.
Line SKU Production Cost	Number	The production cost of the product.

Changelog

Change ID	Reason	Title of Change	Description	# Values	Process Link	Version Link
C1	Data constraint violation.	Fixing mistyping of 'Line SKU Production Cost' Value.	Change Line SKU Production Cost from 432,000 € to 43.2 €.	1	C1-P	C1-V
C2	Data Mistyping.	Fixing mistyping of 'Line Category' value.	Change 'Line Category' from Mini bag to Mini bags.	3	C2-P	C2-V
C3	Blank Data.	Filling Values for 'Order ID'.	Fill Blank Fields to contain the 'Consumer ID' plus the two characters I and J.	2	C3-P	C3-V
C4	Blank Data.	Filling Values for 'Line SKU Production Cost'.	Fill Blank Field of 'MIN3' Line SKU to contain 247.26 €.	1	C4-P	C4-V
C5	Blank Data.	Filling Values for 'Line SKU Production Cost'.	Fill Blank Field of 'MIN4' Line SKU to contain 293.22 €.	1	C5-P	C5-V
C6	Blank Data.	Filling Values for 'Line SKU Production Cost'.	Fill Blank Field of 'MIN2' Line SKU to contain 222.25 €.	1	C6-P	C6-V
C7	Blank Data.	Filling Values for 'Line SKU Production Cost'.	Fill Blank Field of 'MIN8' Line SKU to contain 210.4 €.	1	C7-P	C7-V

C8	Mismatched Data Type.	Converting the data type of “Consumer ID”.	Convert the data type of “Consumer ID” from General to Text.	15	C8-P	C8-V
C9	Mismatched Data Type.	Converting the data type of “Order ID”.	Convert the data type of “Consumer ID” from General to Text.	3678	C9-P	C9-V
C10	Mismatched Data Type.	Converting the data type of “Month”.	Convert the data type of “Month” from General to Text.	3678	C10-P	C10-V
C11	Mismatched Data Type.	Converting the data type of “Year”.	Convert the data type of “Year” from General to Text.	3678	C11-P	C11-V
C12	Mismatched Data Type.	Converting the data type of “Total order value”.	Convert the data type of “Total order value” from Currency to Number.	3678	C12-P	C12-V
C13	Mismatched Data Type.	Converting the data type of “Discount”.	Convert the data type of “Discount” from Currency to Number.	3678	C13-P	C13-V
C14	Mismatched Data Type.	Converting the data type of “Line Value (net discount)”.	Convert the data type of “Line Value (net discount)” from Currency to Number.	3678	C14-P	C14-V
C15	Mismatched Data Type.	Converting the data type of “Line Category”.	Convert the data type of “Line Category” from General to Text.	3678	C15-P	C15-V
C16	Mismatched Data Type.	Converting the data type of “Line SKU”.	Convert the data type of “Line SKU” from General to Text.	3678	C16-P	C16-V
C17	Mismatched Data Type.	Converting the data type of “Line SKU Production Cost”.	Convert the data type of “Line SKU Production Cost” from Currency to Number.	3678	C17-P	C17-V
C18	Adjusting columns for analysis.	Combine the “Month” and “Year” columns.	Combine “Month” and “Year” into “Date”, then remove the original columns.	3678	C18-P	C18-V
C19	Adjusting columns for analysis.	Adding a “Net Profit” column.	Adding a Adding a “Net Profit” column using the formula “Line Value (net discount)” minus “Line SKU Production Cost”.	3678	C19-P	C19-V
C20	Adjusting columns for analysis.	Adding a “Is New Consumer” column.	Adding a “Is New Consumer” column which equal ‘1’ in case	3678	C20-P	C20-V

			of a new customer, '0' otherwise.			
C21	Adjusting columns for analysis.	Adding a "Is Last Order" column.	Adding a "Is Last Order" column which equal '1' in case of a last order made by a customer, '0' otherwise.	3678	C21-P	C21-V
C22	Adjusting columns for analysis.	Adding a "Is Discount" column.	Adding a "Is Discount" column which equal '1' in case of the order is discounted, '0' otherwise.	3678	C22-P	C22-V
C23	Adjusting columns for analysis.	Adding a "Discount Category" column.	Adding a "Discount Category" column which represents the percentage of the order discount out of the total order value.	3678	C23-P	C23-V
C24	Adjusting columns for analysis.	Adding a "Number of Orders" column.	Adding a "Number of Orders" column which represents the total number of orders purchased by a customer.	3678	C24-P	C24-V
C25	Adjusting columns for analysis.	Adding a "Order Number" column.	Adding a "Order Number" column which represents the current order number purchased by a customer.	3678	C25-P	C25-V

Cleaning Process

1. Business Logic (does the data make sense?)

Constrains based on data:

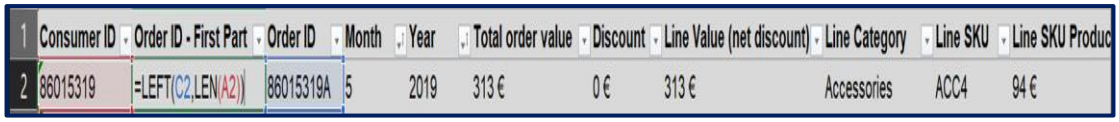
- Order ID = Consumer ID + Alphabetic Char
- Total order value >= Discount
- Line Value (net discount) = Total order value – Discount
- Line SKU Production Cost < Line Value (net discount)

Each constrain was checked using **conditional formatting**:

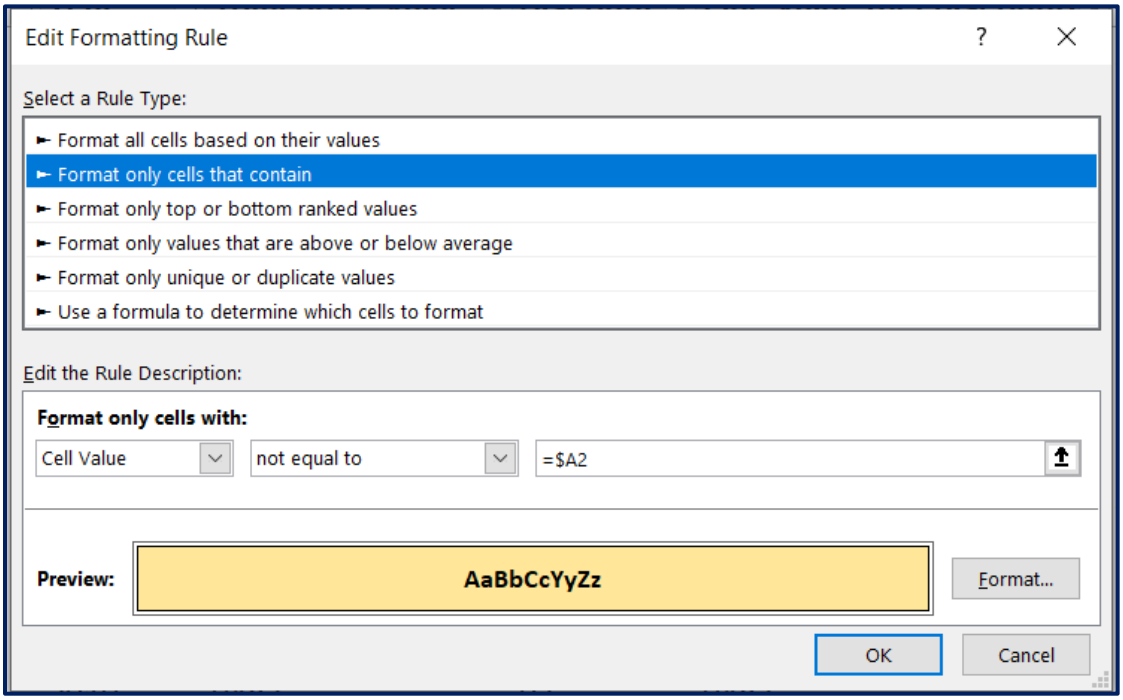
1) Order ID = Consumer ID + Alphabetic Char

To check for this constraint:

1. Add a new column called 'Order ID - First Part'.
2. Extract the first part of the Order ID using both the **LEFT** and **LEN** function.



3. Use **conditional formatting** to check the constraint.



Output:

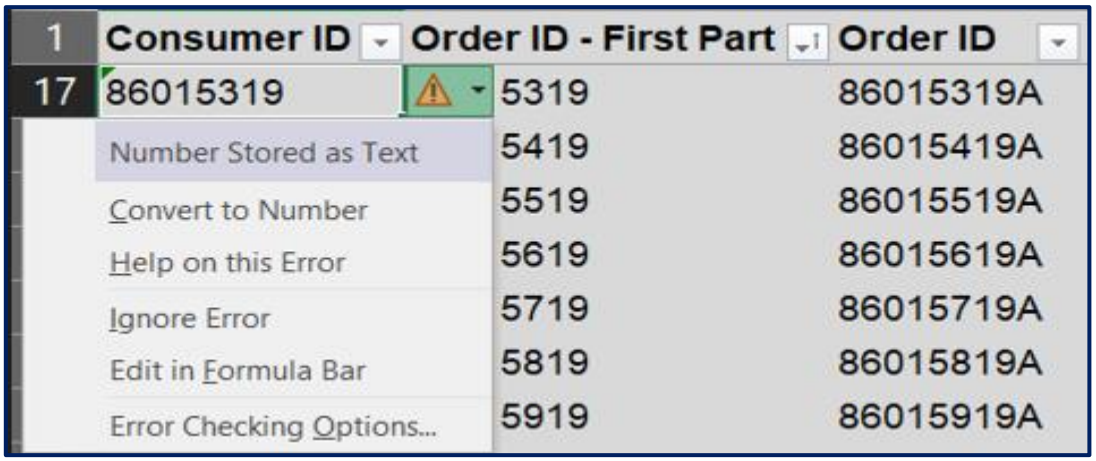
There are 15 values highlighted as a violation, although they don't seem as such.

	Consumer ID	Order ID - First Part	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
2	13134019	13134019	13134019E	2	2020	447 €	0 €	447 €	Apparel	APP3	156 €
3	13121938	13121938	13121938B	3	2020	168 €	0 €	168 €	Small leather goods	SMA4	25 €
4	29022004	29022004	29022004A	4	2020	312 €	47 €	265 €	Sneakers	SNE1	100 €
5	29022008	29022008	29022008A	4	2020	447 €	67 €	380 €	Apparel	APP3	156 €
6	29022012	29022012	29022012A	4	2020	506 €	76 €	430 €	Sandals	SAN12	147 €
7	29022016	29022016	29022016A	4	2020	383 €	57 €	326 €	Apparel	APP1	123 €
8	29022020	29022020	29022020A	4	2020	457 €	69 €	388 €	Shoes	SHO5	123 €
9	29022024	29022024	29022024A	4	2020	447 €	67 €	380 €	Apparel	APP3	156 €
10	29021932	29021932	29021932A	4	2020	383 €	57 €	326 €	Apparel	APP1	123 €
11	29021936	29021936	29021936A	4	2020	210 €	32 €	179 €	Fragrances	FRA2	38 €
12	29021940	29021940	29021940A	4	2020	220 €	33 €	187 €	Fragrances	FRA4	40 €
13	13121938	13121938	13121938C	5	2020	220 €	0 €	220 €	Fragrances	FRA4	40 €
14	13121938	13121938	13121938D	7	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €
15	29021936	29021936	29021936B	8	2020	258 €	0 €	258 €	Accessories	ACC2	72 €
16	13133719	13133719	13133719F	9	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €
17	86015319	86015319	86015319A	5	2019	313 €	0 €	313 €	Accessories	ACC4	94 €

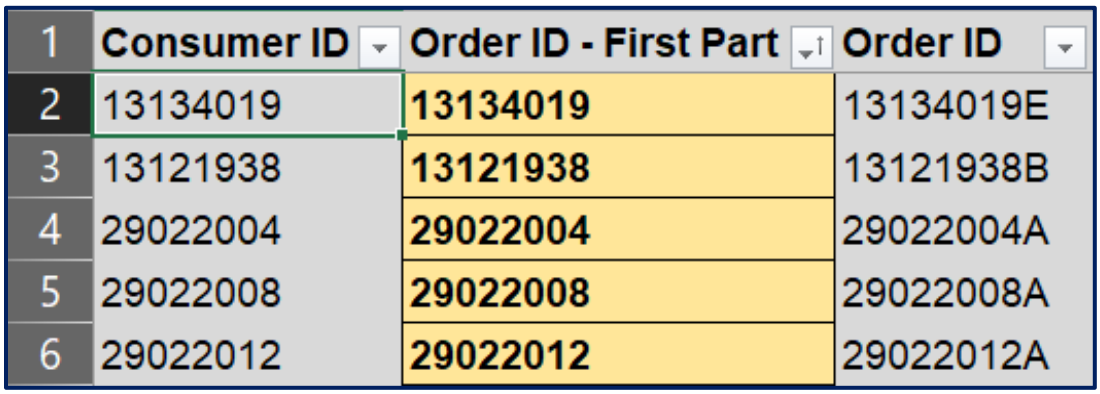
Observation:

A further investigation has found that the highlighted values are because of different data types rather than violating the data constraint.

- "Consumer ID" that <u>didn't</u> trigger a highlight in the "Order ID – First Part" was stored as text.



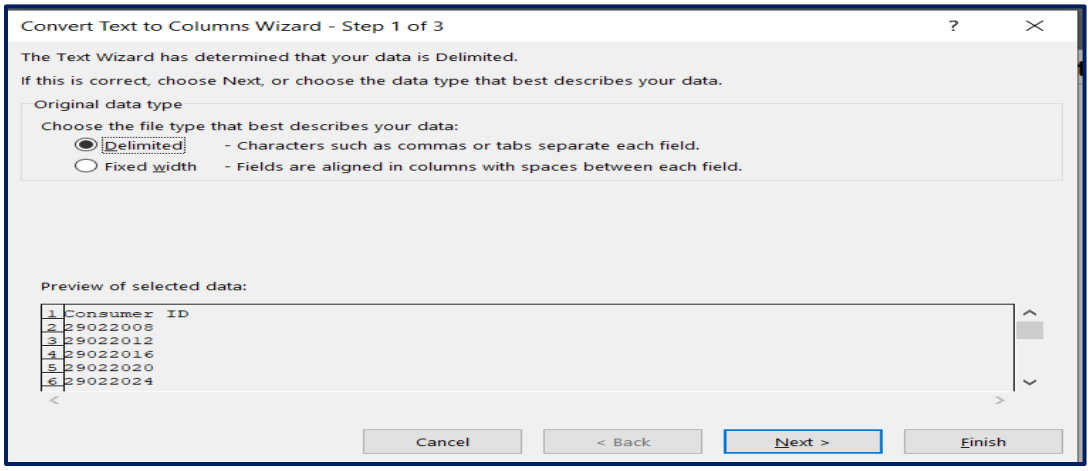
- "Consumer ID" that <u>did</u> trigger a highlight in the "Order ID – First Part" wasn't stored as text.

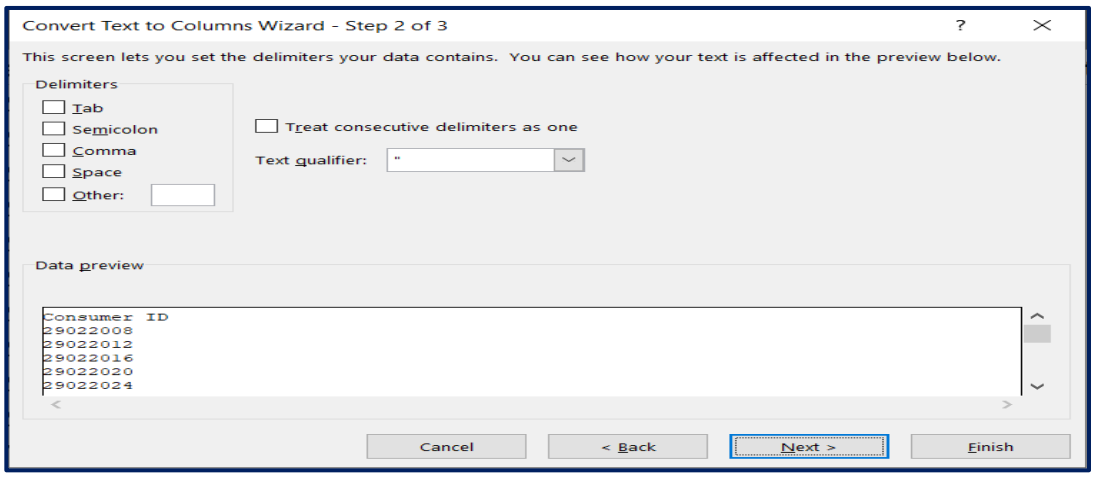


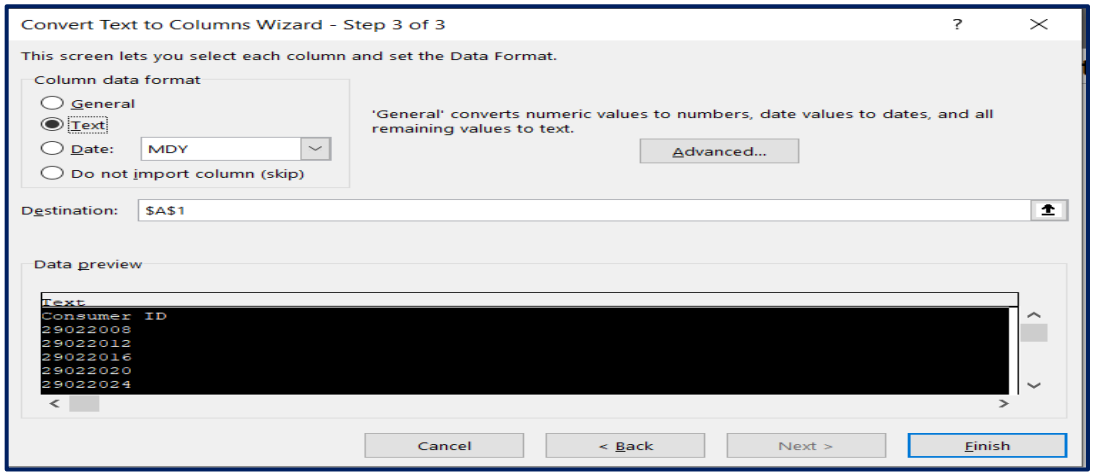
Correctness:

Steps:

1. Convert the data type of the “Consumer ID” from General to Text using **Text to Columns** feature.





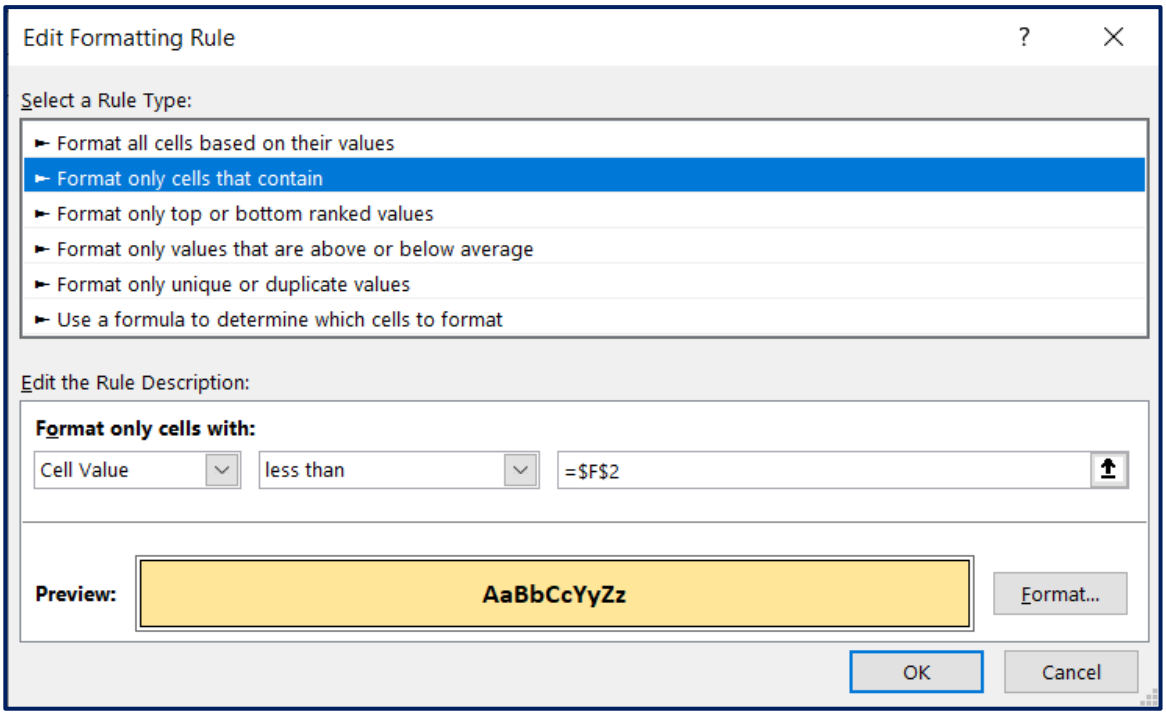


NOTE:

- The Text to Columns feature was used rather than converting the type simply from the home tap, because it would require re-entering the data again.
- This change occurred as the 8th change.

	Consumer ID	Order ID - First Part	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
2	29022008	29022008	29022008A	4	2020	447 €	67 €	380 €	Apparel	APP3	156 €
3	29022012	29022012	29022012A	4	2020	506 €	76 €	430 €	Sandals	SAN12	147 €
4	29022016	29022016	29022016A	4	2020	383 €	57 €	326 €	Apparel	APP1	123 €
5	29022020	29022020	29022020A	4	2020	457 €	69 €	388 €	Shoes	SHO5	123 €
6	29022024	29022024	29022024A	4	2020	447 €	67 €	380 €	Apparel	APP3	156 €
7	29021932	29021932	29021932A	4	2020	383 €	57 €	326 €	Apparel	APP1	123 €
8	29021936	29021936	29021936A	4	2020	210 €	32 €	179 €	Fragrances	FRA2	38 €
9	29021940	29021940	29021940A	4	2020	220 €	33 €	187 €	Fragrances	FRA4	40 €
10	13121938	13121938	13121938C	5	2020	220 €	0 €	220 €	Fragrances	FRA4	40 €
11	13121938	13121938	13121938D	7	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €
12	20170319	20170319	20170319B	12	2020	383 €	0 €	383 €	Apparel	APP1	123 €
13	13134019	13134019	13134019E	2	2020	447 €	0 €	447 €	Apparel	APP3	156 €
14	13121938	13121938	13121938B	3	2020	168 €	0 €	168 €	Small leather goods	SMA4	25 €
15	29022004	29022004	29022004A	4	2020	312 €	47 €	265 €	Sneakers	SNE1	100 €
16	29021936	29021936	29021936B	8	2020	258 €	0 €	258 €	Accessories	ACC2	72 €
17	13133719	13133719	13133719F	9	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €

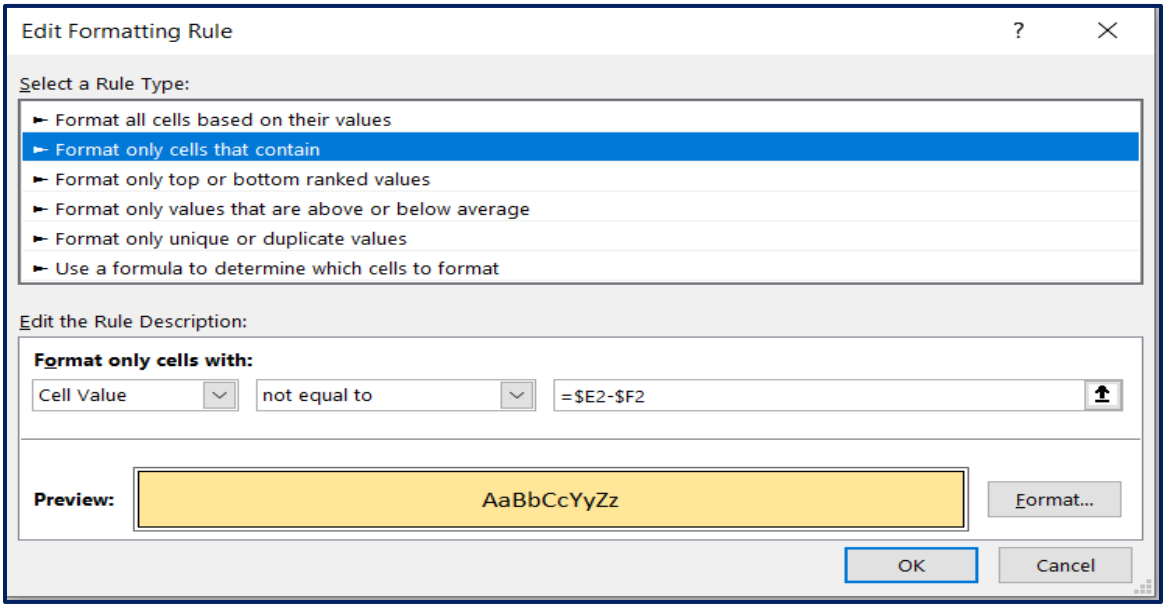
2) Total order value >= Discount



Output:

No value violates the constraint.

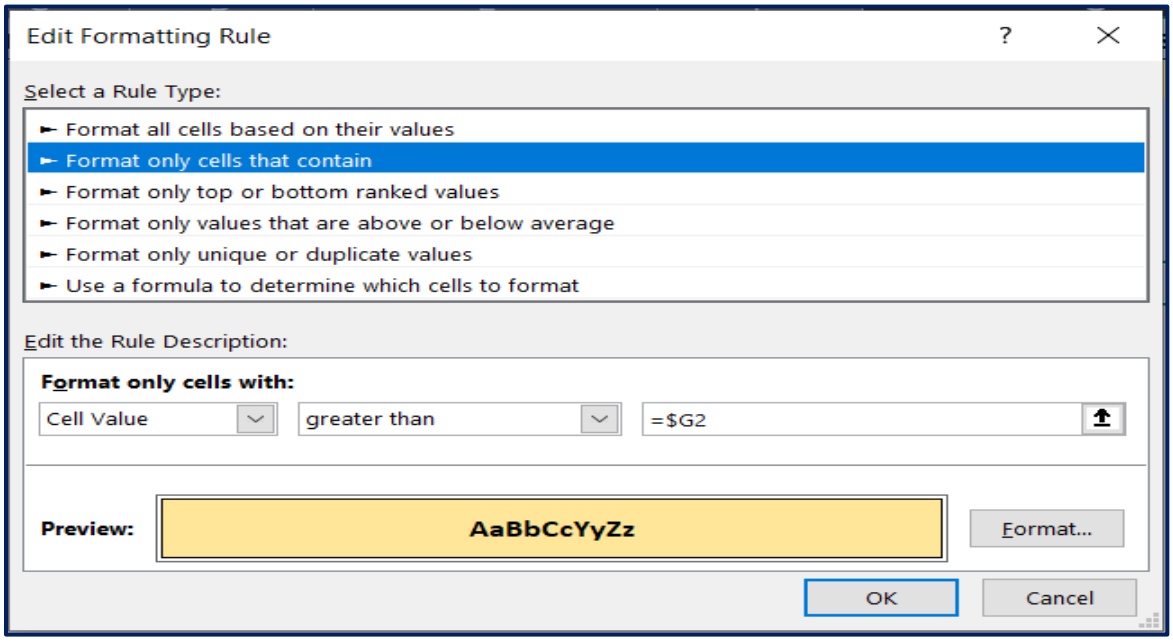
3) Line Value (net discount) = Total order value – Discount



Output:

No value violates the constraint.

4) Line SKU Production Cost < Line Value (net discount)



Output:
one value violates the constraint

1	Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
2	13133719	13133719F	9	2020	240 €	0 €	240 €	Fragrances	FRA5	432,000 €

Correctness:

Steps:

1. Filter for the 'FRA5' Line SKU

1	Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
2	13133719	13133719F	9	2020	240 €	0 €	240 €	Fragrances	FRA5	432,000 €
5	13121938	13121938D	7	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €
33	13132719	13132719A	1	2020	240 €	48 €	192 €	Fragrances	FRA5	43 €
35	13132719	13132719C	5	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €
37	13132719	13132719E	7	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €
44	13132819	13132819E	6	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €
46	13132919	13132919A	1	2020	240 €	48 €	192 €	Fragrances	FRA5	43 €
68	13133519	13133519E	9	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €
85	13133819	13133819B	4	2020	240 €	36 €	204 €	Fragrances	FRA5	43 €
167	13131719	13131719C	10	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €

Observations:

170 records have the category 'FRA5'. All these records have a Line SKU Production Cost of 43.2 €.

2. Change the Line SKU Production Cost of the first record from 432,000 € to 43.2 €.

1	Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
2	13133719	13133719F	9	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €
5	13121938	13121938D	7	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €
33	13132719	13132719A	1	2020	240 €	48 €	192 €	Fragrances	FRA5	43 €
35	13132719	13132719C	5	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €
37	13132719	13132719E	7	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €
44	13132819	13132819E	6	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €
46	13132919	13132919A	1	2020	240 €	48 €	192 €	Fragrances	FRA5	43 €
68	13133519	13133519E	9	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €
85	13133819	13133819B	4	2020	240 €	36 €	204 €	Fragrances	FRA5	43 €
167	13131719	13131719C	10	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €

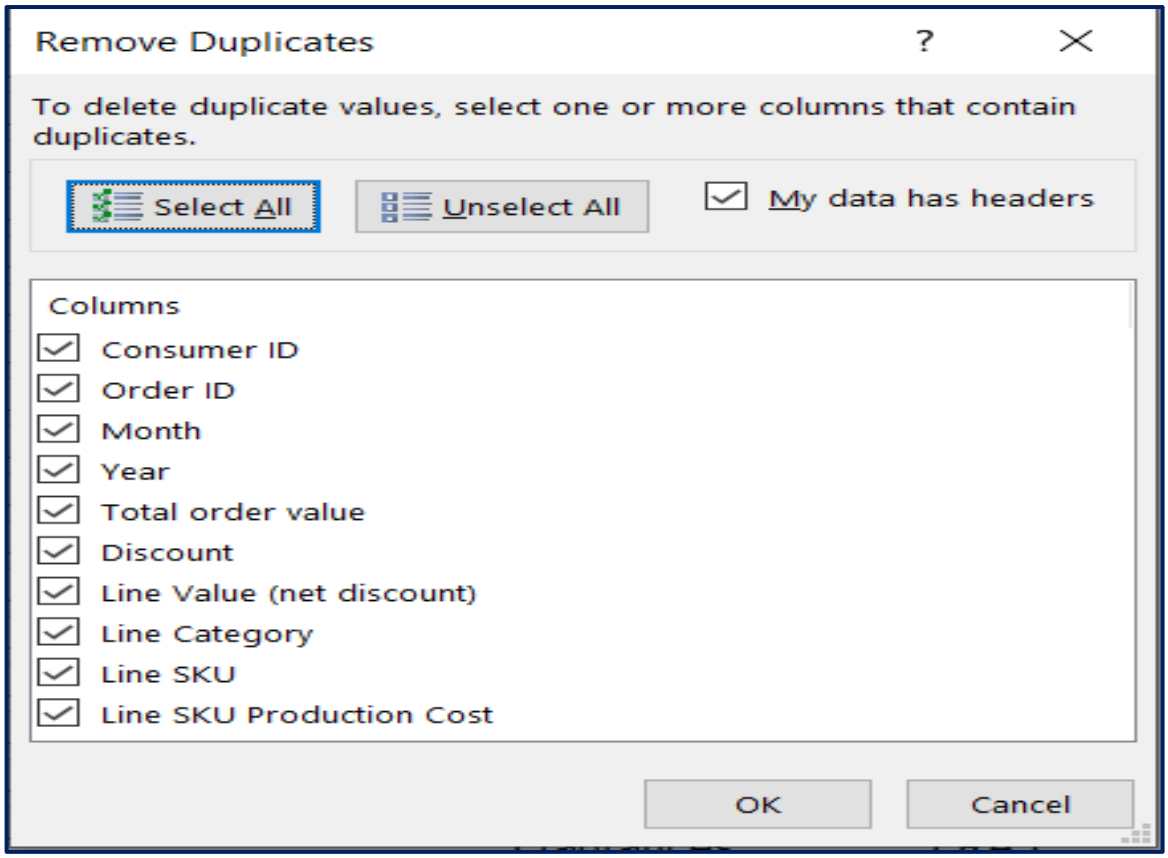
2. Check for Duplicates

Unique Constraints:

- A Record can't be duplicated.
- Order ID can't appear twice.

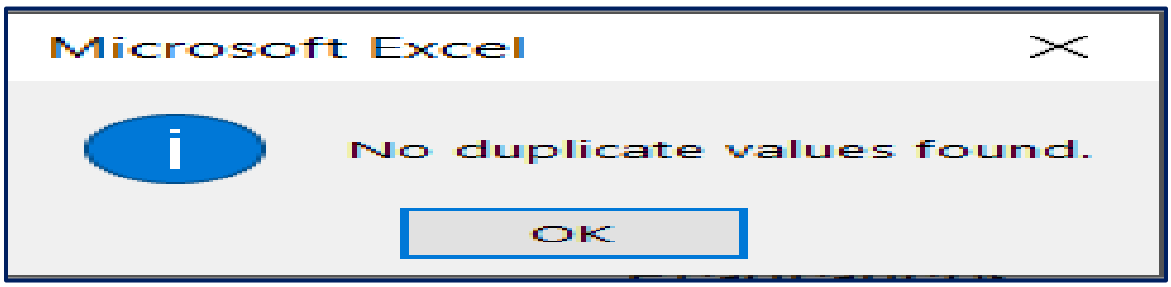
Check for each constrain using **Remove Duplicates** and **Conditional Formatting**:

1) A Record can't be duplicated.
   use **Remove Duplicates** to find and remove duplicate records.



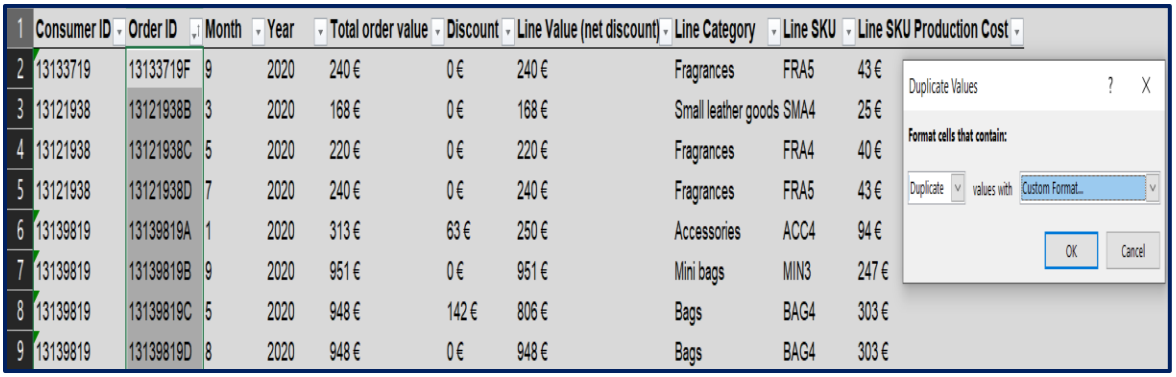
Output:

No record violates the constraint.



2) Order ID can't appear twice.

Check for duplicates using **conditional formatting**:



Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
13133719	13133719F	9	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €
13121938	13121938B	3	2020	168 €	0 €	168 €	Small leather goods	SMA4	25 €
13121938	13121938C	5	2020	220 €	0 €	220 €	Fragrances	FRA4	40 €
13121938	13121938D	7	2020	240 €	0 €	240 €	Fragrances	FRA5	43 €
13139819	13139819A	1	2020	313 €	63 €	250 €	Accessories	ACC4	94 €
13139819	13139819B	9	2020	951 €	0 €	951 €	Mini bags	MIN3	247 €
13139819	13139819C	5	2020	948 €	142 €	806 €	Bags	BAG4	303 €
13139819	13139819D	8	2020	948 €	0 €	948 €	Bags	BAG4	303 €

Output:

Two values violate the constraint.

Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
13134019	#N/A	3	2020	447 €	67 €	380 €	Apparel	APP3	156 €
13134019	#N/A	3	2020	345 €	52 €	293 €	Apparel	APP2	107 €

as the value of the Order ID is N/A, no values will be removed.

3. Check for Mistyping

In the 'Line Category' column there are two categories with the same name:

- Mini bags
- Mini bag

but one is plural and the other is singular.

	Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
929	23236919	23236919A	7	2019	951 €	0 €	951 €	Mini bags	MIN3	247 €
934	23237819	23237819A	7	2019	526 €	0 €	526 €	Mini bags	MIN8	210 €
942	23239819	23239819A	7	2019	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
943	23239919	23239919A	7	2019	951 €	0 €	951 €	Mini bags	MIN3	247 €
945	13995219	13995219A	7	2019	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
1036	86014519	86014519A	6	2019	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
1040	86015419	86015419A	5	2019	889 €	0 €	889 €	Mini bags	MIN2	222 €
1041	86015519	86015519A	5	2019	474 €	0 €	474 €	Mini bags	MIN9	175 €
1049	86018019	86018019A	5	2019	951 €	0 €	951 €	Mini bag	MIN3	247 €

Correctness:

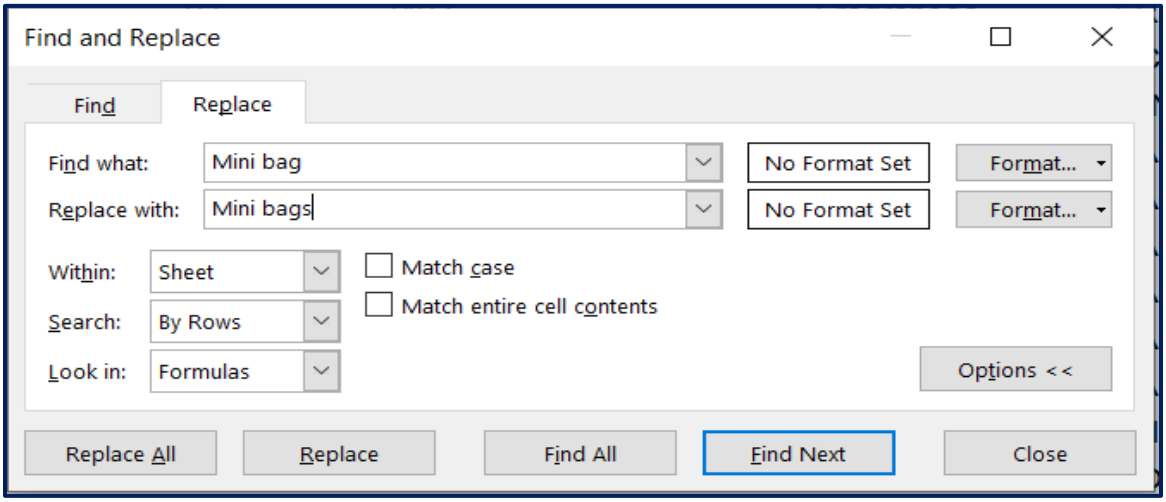
As the two 'Line Category' have records with the same 'Line SKU'. Then the two are likely the same.

steps:

1. Filter for the 'Mini bag' Line Category.

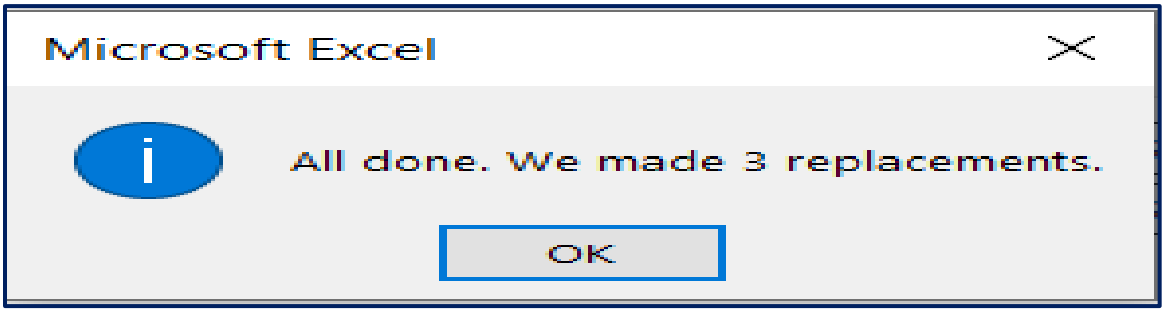
	Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
1049	86018019	86018019A	5	2019	951 €	0 €	951 €	Mini bag	MIN3	247 €
1224	96010919	96010919A	9	2020	951 €	0 €	951 €	Mini bag	MIN3	247 €
1276	15156619	15156619A	9	2020	992 €	0 €	992 €	Mini bag	MIN1	278 €

2. Change 'Mini bag' to 'Mini bags' using **Find and Replace**.

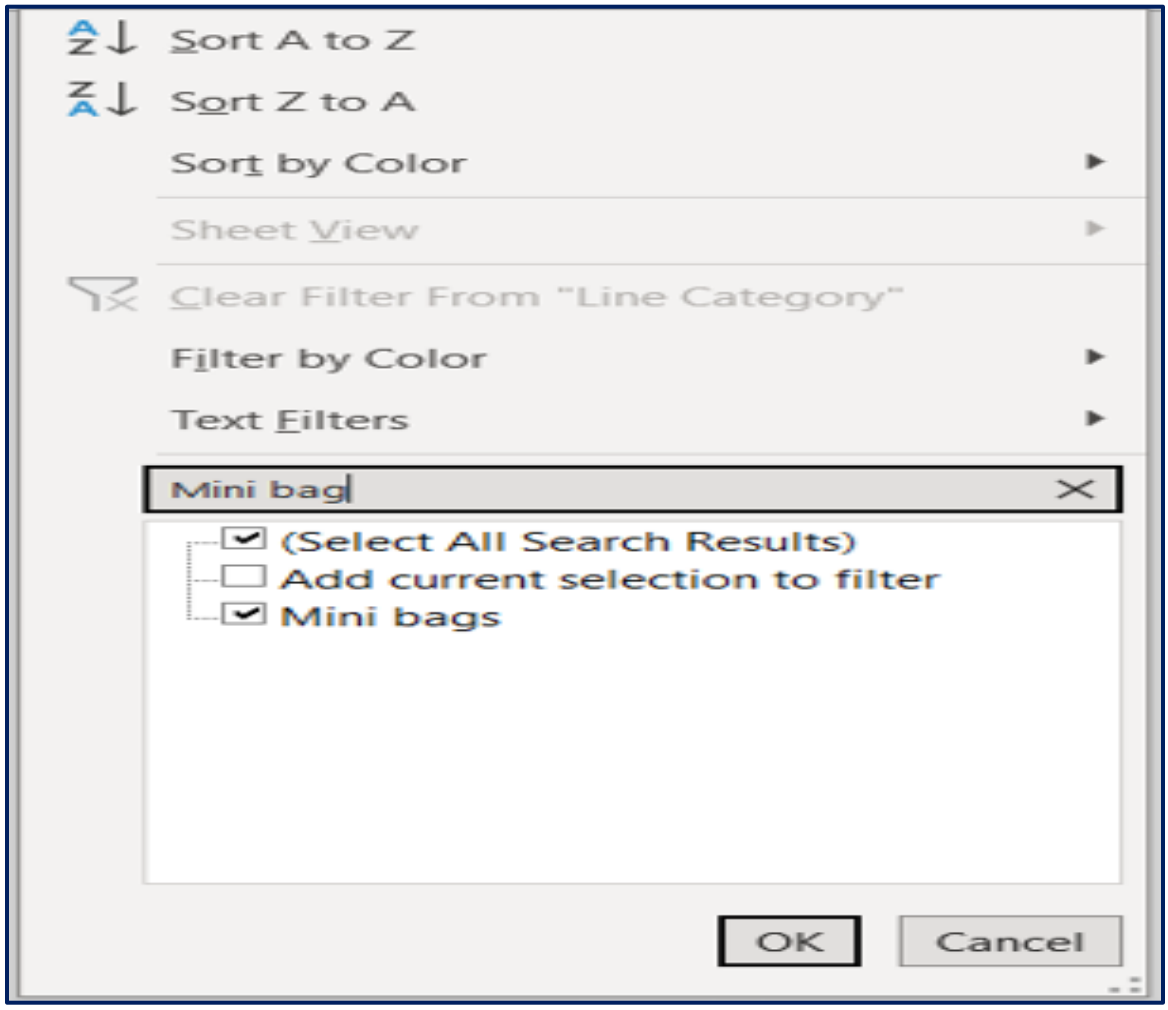


Output:

Three values were changed.



3. Validate using the Filter Menu.



4. Check for Blank Fields

There are blanks fields in:

- Order ID
- Line SKU Production Cost

Let's investigate each column:

1) Order ID
   Use Filtration to find Blank fields.

	Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
3678	13134019	#N/A	3	2020	447 €	67 €	380 €	Apparel	APP3	156 €
3679	13134019	#N/A	3	2020	345 €	52 €	293 €	Apparel	APP2	107 €

<u>Correctness</u>:

Steps:

1. As the blanks field were for the 'Consumer ID' => 13134019, Filter only this one.

	Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
92	13134019	13134019A	1	2020	478 €	96 €	382 €	Sandals	SAN8	120 €
93	13134019	13134019B	2	2020	267 €	0 €	267 €	Sneakers	SNE6	91 €
94	13134019	13134019C	3	2020	383 €	57 €	326 €	Apparel	APP1	123 €
95	13134019	13134019D	2	2020	383 €	0 €	383 €	Apparel	APP1	123 €
96	13134019	13134019E	2	2020	447 €	0 €	447 €	Apparel	APP3	156 €
97	13134019	13134019F	3	2020	447 €	67 €	380 €	Apparel	APP3	156 €
98	13134019	13134019G	4	2020	345 €	52 €	293 €	Apparel	APP2	107 €
99	13134019	13134019H	3	2020	345 €	52 €	293 €	Apparel	APP2	107 €
3678	13134019	#N/A	3	2020	447 €	67 €	380 €	Apparel	APP3	156 €
3679	13134019	#N/A	3	2020	345 €	52 €	293 €	Apparel	APP2	107 €

2. Change the two blank fields to contain the Consumer ID plus the two characters I and J.

	Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
92	13134019	13134019A	1	2020	478 €	96 €	382 €	Sandals	SAN8	120 €
93	13134019	13134019B	2	2020	267 €	0 €	267 €	Sneakers	SNE6	91 €
94	13134019	13134019C	3	2020	383 €	57 €	326 €	Apparel	APP1	123 €
95	13134019	13134019D	2	2020	383 €	0 €	383 €	Apparel	APP1	123 €
96	13134019	13134019E	2	2020	447 €	0 €	447 €	Apparel	APP3	156 €
97	13134019	13134019F	3	2020	447 €	67 €	380 €	Apparel	APP3	156 €
98	13134019	13134019G	4	2020	345 €	52 €	293 €	Apparel	APP2	107 €
99	13134019	13134019H	3	2020	345 €	52 €	293 €	Apparel	APP2	107 €
3678	13134019	13134019I	3	2020	447 €	67 €	380 €	Apparel	APP3	156 €
3679	13134019	13134019J	3	2020	345 €	52 €	293 €	Apparel	APP2	107 €

2) Line SKU Production Cost

Use Filtration to find Blank fields.

	Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
229	48012919	48012919A	1	2020	951 €	190 €	761 €	Mini bags	MIN3	
263	48015719	48015719A	12	2019	1,086 €	217 €	869 €	Mini bags	MIN4	
1644	88004419	88004419A	8	2020	889 €	0 €	889 €	Mini bags	MIN2	
1986	59014619	59014619A	6	2020	526 €	0 €	526 €	Mini bags	MIN8	

<u>Correctness</u>:

Steps:

1. Filter for 'Line SKU' equal MIN3.

	Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
7	13139819	13139819B	9	2020	951 €	0 €	951 €	Mini bags	MIN3	247 €
229	48012919	48012919A	1	2020	951 €	190 €	761 €	Mini bags	MIN3	
243	48013819	48013819A	1	2020	951 €	190 €	761 €	Mini bags	MIN3	247 €
567	45502919	45502919A	10	2019	951 €	0 €	951 €	Mini bags	MIN3	247 €
929	23236919	23236919A	7	2019	951 €	0 €	951 €	Mini bags	MIN3	247 €
943	23239919	23239919A	7	2019	951 €	0 €	951 €	Mini bags	MIN3	247 €
1049	86018019	86018019A	5	2019	951 €	0 €	951 €	Mini bags	MIN3	247 €
1211	94051619	94051619A	9	2020	951 €	0 €	951 €	Mini bags	MIN3	247 €
1224	96010919	96010919A	9	2020	951 €	0 €	951 €	Mini bags	MIN3	247 €
1269	15154919	15154919A	9	2020	951 €	0 €	951 €	Mini bags	MIN3	247 €
1824	79011719	79011719A	7	2020	951 €	0 €	951 €	Mini bags	MIN3	247 €

Observations:

27 records have the Line SKU 'MIN3'. All these records have a Line SKU Production Cost of 247.26 €.

Fill the 'Line SKU Production Cost' of the 'MIN3' to 247.26 €.

	Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
7	13139819	13139819B	9	2020	951 €	0 €	951 €	Mini bags	MIN3	247 €
229	48012919	48012919A	1	2020	951 €	190 €	761 €	Mini bags	MIN3	247 €
243	48013819	48013819A	1	2020	951 €	190 €	761 €	Mini bags	MIN3	247 €
567	45502919	45502919A	10	2019	951 €	0 €	951 €	Mini bags	MIN3	247 €
929	23236919	23236919A	7	2019	951 €	0 €	951 €	Mini bags	MIN3	247 €
943	23239919	23239919A	7	2019	951 €	0 €	951 €	Mini bags	MIN3	247 €
1049	86018019	86018019A	5	2019	951 €	0 €	951 €	Mini bags	MIN3	247 €
1211	94051619	94051619A	9	2020	951 €	0 €	951 €	Mini bags	MIN3	247 €
1224	96010919	96010919A	9	2020	951 €	0 €	951 €	Mini bags	MIN3	247 €
1269	15154919	15154919A	9	2020	951 €	0 €	951 €	Mini bags	MIN3	247 €
1824	79011719	79011719A	7	2020	951 €	0 €	951 €	Mini bags	MIN3	247 €

2. Filter for ‘Line SKU’ equal MIN4.

1	Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
263	48015719	48015719A	12	2019	1,086 €	217 €	869 €	Mini bags	MIN4	
576	45503819	45503819A	10	2019	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
707	45205819	45205819A	8	2019	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
942	23239819	23239819A	7	2019	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
945	13995219	13995219A	7	2019	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
1036	86014519	86014519A	6	2019	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
1054	99052119	99052119A	5	2019	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
1182	94030819	94030819A	9	2020	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
1197	94040719	94040719A	9	2020	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
1420	24243919	24243919A	9	2020	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
1557	56014819	56014819A	8	2020	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €

Observations:

33 records have the Line SKU ‘MIN4’. All these records have a Line SKU Production Cost of 293.22 €.

Fill the ‘Line SKU Production Cost’ of the ‘MIN4’ to 293.22 €.

1	Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
263	48015719	48015719A	12	2019	1,086 €	217 €	869 €	Mini bags	MIN4	293 €
576	45503819	45503819A	10	2019	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
707	45205819	45205819A	8	2019	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
942	23239819	23239819A	7	2019	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
945	13995219	13995219A	7	2019	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
1036	86014519	86014519A	6	2019	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
1054	99052119	99052119A	5	2019	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
1182	94030819	94030819A	9	2020	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
1197	94040719	94040719A	9	2020	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
1420	24243919	24243919A	9	2020	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €
1557	56014819	56014819A	8	2020	1,086 €	0 €	1,086 €	Mini bags	MIN4	293 €

3. Filter for ‘Line SKU’ equal MIN2.

1	Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
175	13992219	13992219A	1	2020	889 €	178 €	711 €	Mini bags	MIN2	222 €
282	99060219	99060219A	12	2019	889 €	178 €	711 €	Mini bags	MIN2	222 €
649	45070819	45070819A	9	2019	889 €	0 €	889 €	Mini bags	MIN2	222 €
650	45070919	45070919A	9	2019	889 €	0 €	889 €	Mini bags	MIN2	222 €
717	45206819	45206819A	8	2019	889 €	0 €	889 €	Mini bags	MIN2	222 €
725	45207619	45207619A	8	2019	889 €	0 €	889 €	Mini bags	MIN2	222 €
726	45207719	45207719A	8	2019	889 €	0 €	889 €	Mini bags	MIN2	222 €
1040	86015419	86015419A	5	2019	889 €	0 €	889 €	Mini bags	MIN2	222 €
1068	99056019	99056019A	5	2019	889 €	0 €	889 €	Mini bags	MIN2	222 €
1126	99030319	99030319A	9	2020	889 €	0 €	889 €	Mini bags	MIN2	222 €
1385	99020819	99020819A	9	2020	889 €	0 €	889 €	Mini bags	MIN2	222 €
1566	56015719	56015719A	8	2020	889 €	0 €	889 €	Mini bags	MIN2	222 €
1575	68010919	68010919A	8	2020	889 €	0 €	889 €	Mini bags	MIN2	222 €
1617	88001719	88001719A	8	2020	889 €	0 €	889 €	Mini bags	MIN2	222 €
1627	88002719	88002719A	8	2020	889 €	0 €	889 €	Mini bags	MIN2	222 €
1644	88004419	88004419A	8	2020	889 €	0 €	889 €	Mini bags	MIN2	

Observations:

29 records have the Line SKU ‘MIN2’. All these records have a Line SKU Production Cost of 222.25 €.

Fill the ‘Line SKU Production Cost’ of the ‘MIN2’ to 222.25 €.

1	Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
282	99060219	99060219A	12	2019	889 €	178 €	711 €	Mini bags	MIN2	222 €
649	45070819	45070819A	9	2019	889 €	0 €	889 €	Mini bags	MIN2	222 €
650	45070919	45070919A	9	2019	889 €	0 €	889 €	Mini bags	MIN2	222 €
717	45206819	45206819A	8	2019	889 €	0 €	889 €	Mini bags	MIN2	222 €
725	45207619	45207619A	8	2019	889 €	0 €	889 €	Mini bags	MIN2	222 €
726	45207719	45207719A	8	2019	889 €	0 €	889 €	Mini bags	MIN2	222 €
1040	86015419	86015419A	5	2019	889 €	0 €	889 €	Mini bags	MIN2	222 €
1068	99056019	99056019A	5	2019	889 €	0 €	889 €	Mini bags	MIN2	222 €
1126	99030319	99030319A	9	2020	889 €	0 €	889 €	Mini bags	MIN2	222 €
1385	99020819	99020819A	9	2020	889 €	0 €	889 €	Mini bags	MIN2	222 €
1566	56015719	56015719A	8	2020	889 €	0 €	889 €	Mini bags	MIN2	222 €
1575	68010919	68010919A	8	2020	889 €	0 €	889 €	Mini bags	MIN2	222 €
1617	88001719	88001719A	8	2020	889 €	0 €	889 €	Mini bags	MIN2	222 €
1627	88002719	88002719A	8	2020	889 €	0 €	889 €	Mini bags	MIN2	222 €
1644	88004419	88004419A	8	2020	889 €	0 €	889 €	Mini bags	MIN2	222 €

4. Filter for ‘Line SKU’ equal MIN8.

1	Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
566	45502819	45502819A	10	2019	526 €	0 €	526 €	Mini bags	MIN8	210 €
934	23237819	23237819A	7	2019	526 €	0 €	526 €	Mini bags	MIN8	210 €
1050	86018119	86018119A	5	2019	526 €	0 €	526 €	Mini bags	MIN8	210 €
1198	94040819	94040819A	9	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €
1662	88006419	88006419A	8	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €
1832	79012519	79012519A	7	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €
1909	79020719	79020719A	7	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €
1986	59014619	59014619A	6	2020	526 €	0 €	526 €	Mini bags	MIN8	
2018	59020919	59020919A	6	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €
2938	12790167	12790167B	10	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €
2940	12790167	12790167D	11	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €
2944	12790169	12790169A	3	2020	526 €	79 €	447 €	Mini bags	MIN8	210 €
2960	12790177	12790177B	12	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €
3597	62034819	62034819F	10	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €

Observations:

17 records have the Line SKU ‘MIN8’. All these records have a Line SKU Production Cost of 210.4 €.

Fill the ‘Line SKU Production Cost’ of the ‘MIN8’ to 210.4 €.

1	Consumer ID	Order ID	Month	Year	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost
566	45502819	45502819A	10	2019	526 €	0 €	526 €	Mini bags	MIN8	210 €
934	23237819	23237819A	7	2019	526 €	0 €	526 €	Mini bags	MIN8	210 €
1050	86018119	86018119A	5	2019	526 €	0 €	526 €	Mini bags	MIN8	210 €
1198	94040819	94040819A	9	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €
1662	88006419	88006419A	8	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €
1832	79012519	79012519A	7	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €
1909	79020719	79020719A	7	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €
1986	59014619	59014619A	6	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €
2018	59020919	59020919A	6	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €
2938	12790167	12790167B	10	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €
2940	12790167	12790167D	11	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €
2944	12790169	12790169A	3	2020	526 €	79 €	447 €	Mini bags	MIN8	210 €
2960	12790177	12790177B	12	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €
3597	62034819	62034819F	10	2020	526 €	0 €	526 €	Mini bags	MIN8	210 €

5. Check for mismatched data types

From the Metadata:

- Consumer ID => Text
- Order ID => Text
- Month => Text
- Year => Text
- Total order value => Number
- Discount =>Number
- Line Value (net discount) => Number
- Line Category => Text
- Line SKU => Text
- Line SKU Production Cost => Number

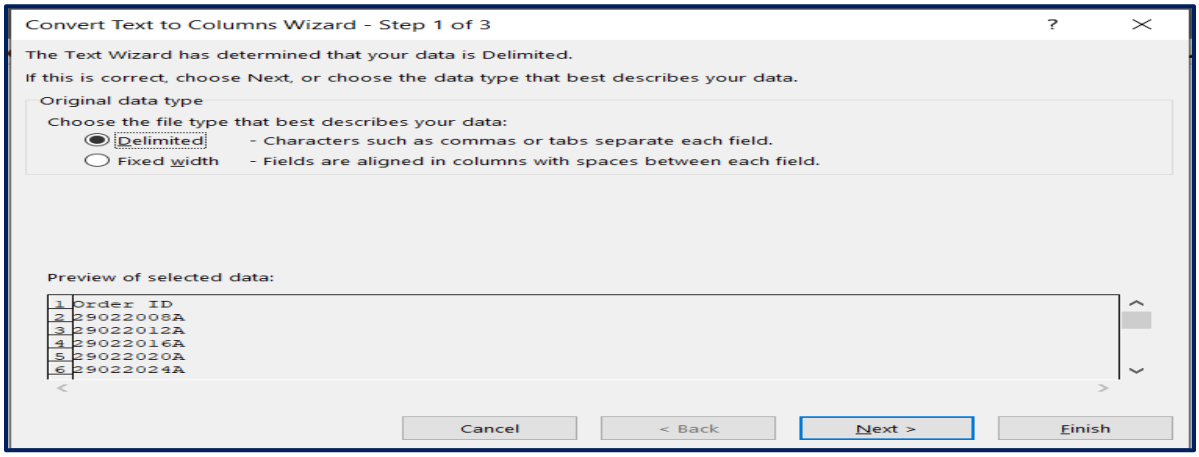
As only the Consumer ID was in the right data type, the rest column's data types will be converted.

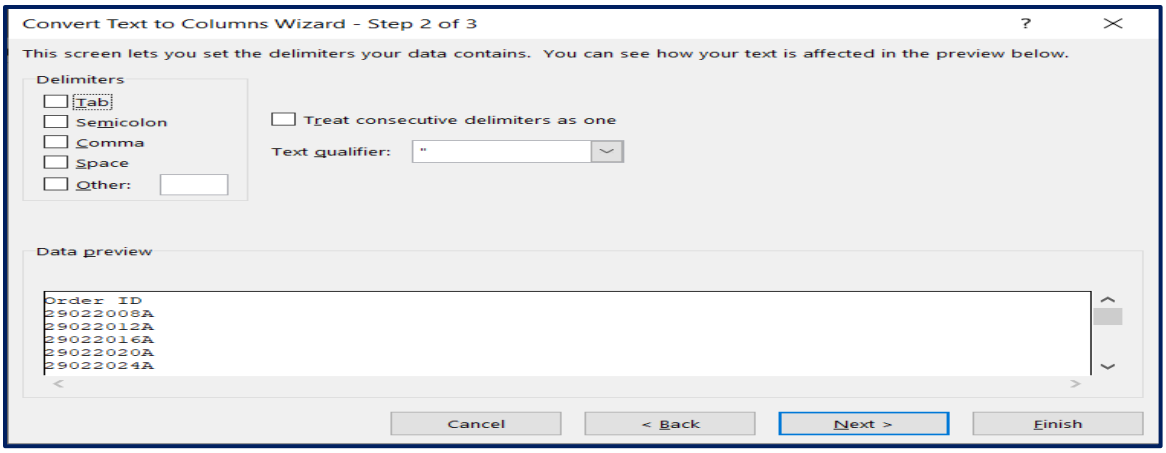
Note:

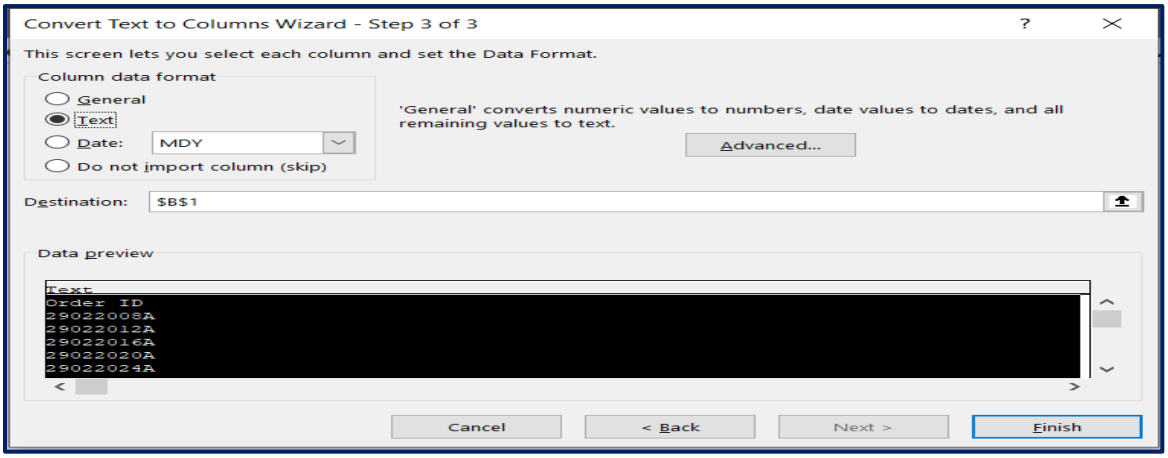
1. Order ID => Text

This column will be converted from General to Text using the **Text to Columns** feature.

NOTE: The Text to Columns feature was used rather than converting the type simply from the home tap, because it would require re-entering the data again.



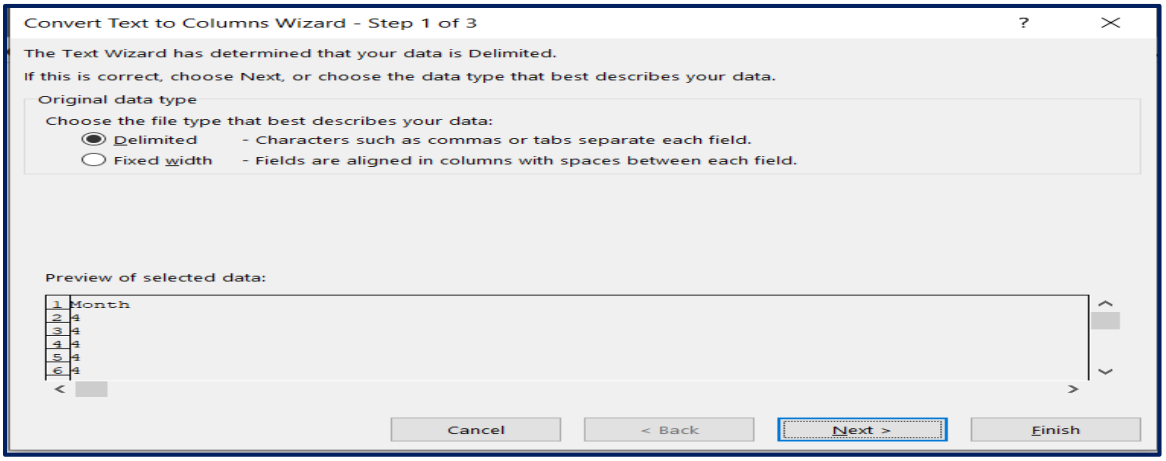


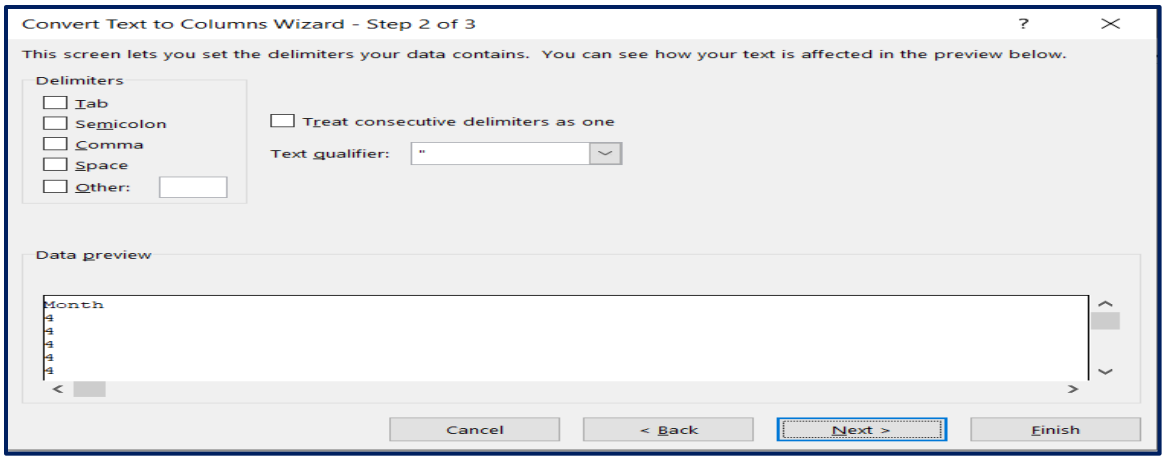


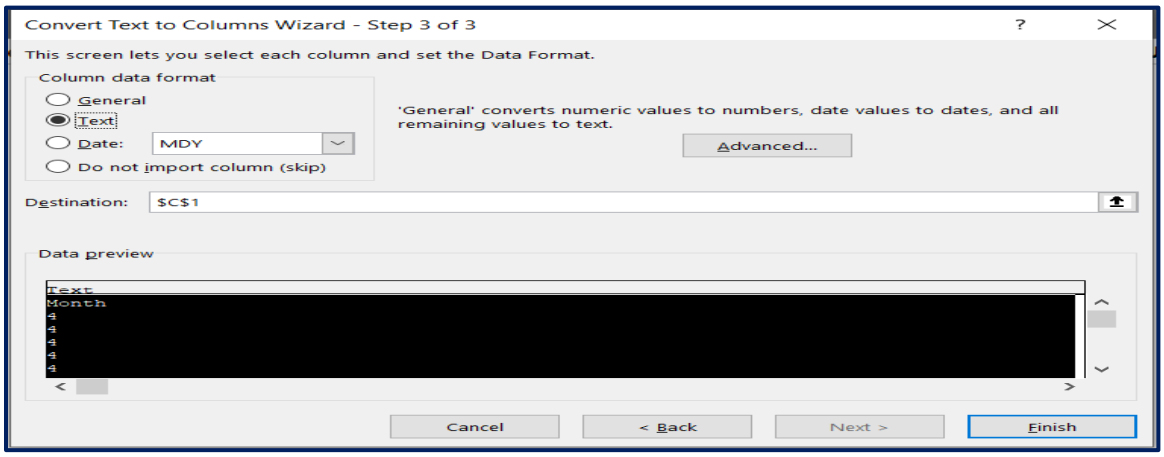
2. Month => Text

This column will be converted from General to Text using the **Text to Columns** feature.

NOTE: The Text to Columns feature was used rather than converting the type simply from the home tap, because it would require re-entering the data again.



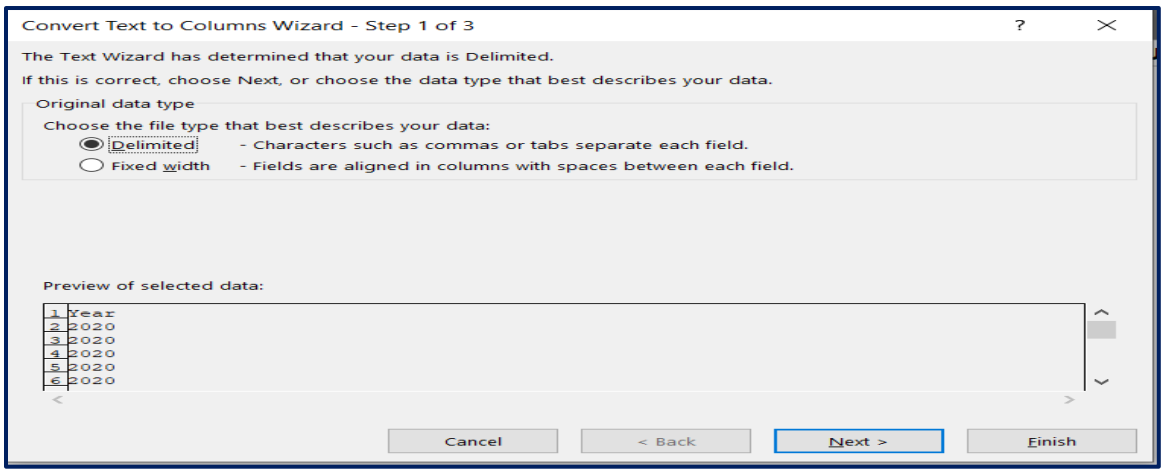


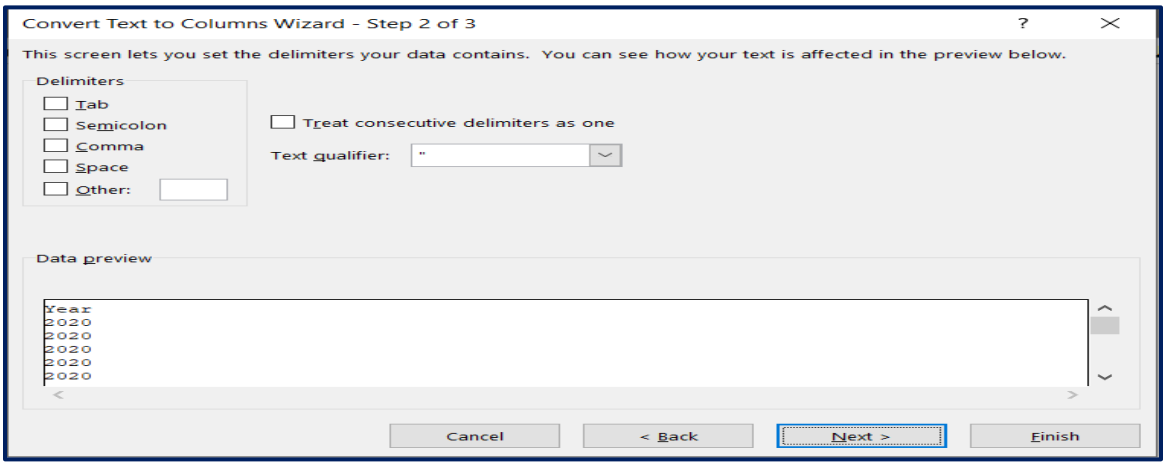


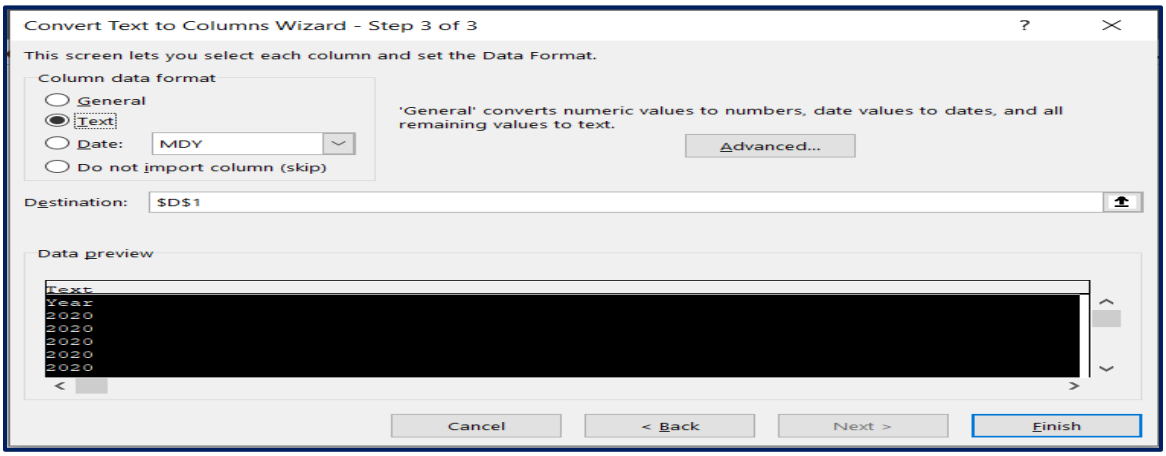
3. Year => Text

This column will be converted from General to Text using the **Text to Columns** feature.

NOTE: The Text to Columns feature was used rather than converting the type simply from the home tap, because it would require re-entering the data again.

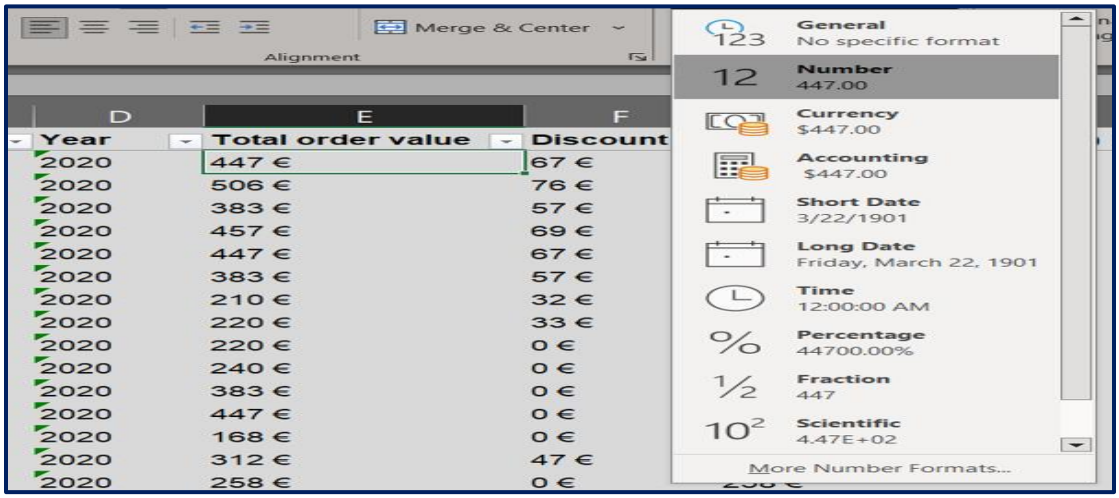




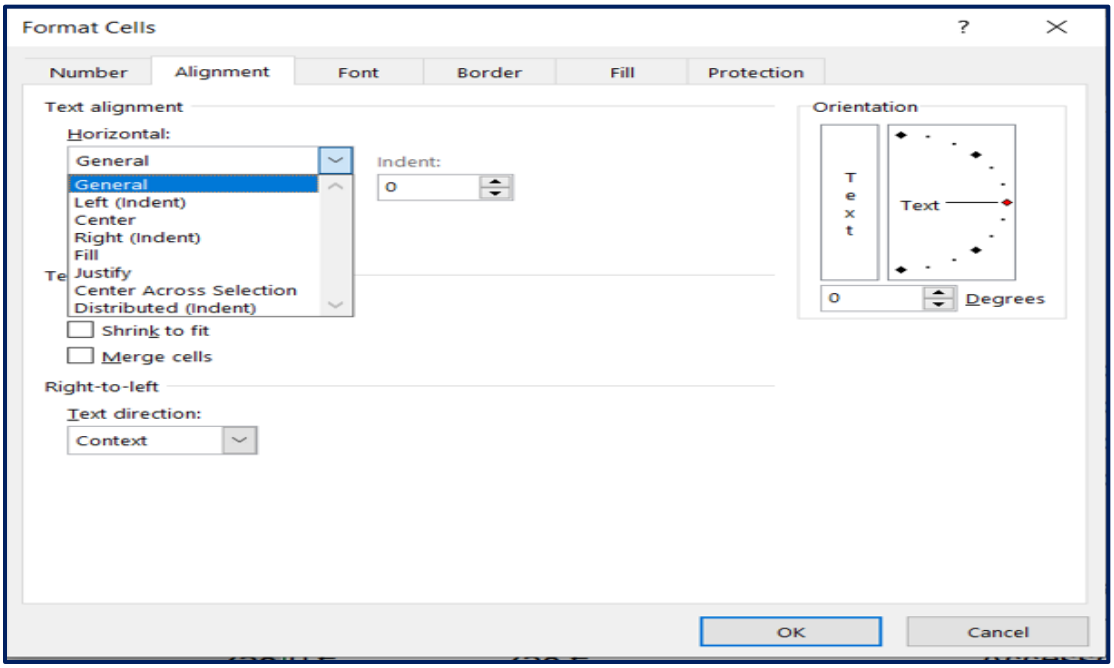


3. Total order value => Number
   This column will be converted from Currency to Number.

   Steps:
   1. convert from currency to Number from the home tap.



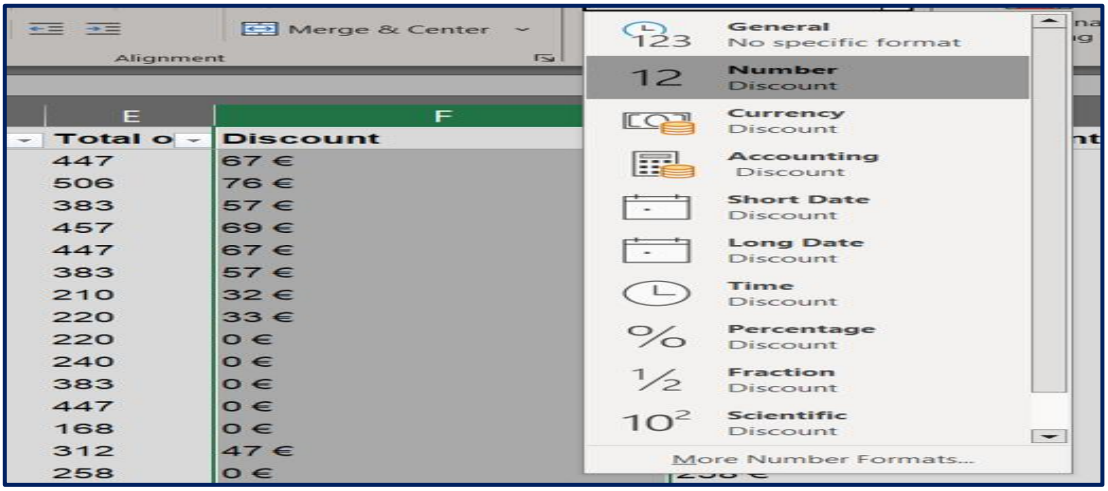
2. As the numbers are still aligned at left, to make it right aligned as the standard, we will make the alignment general instead of right.



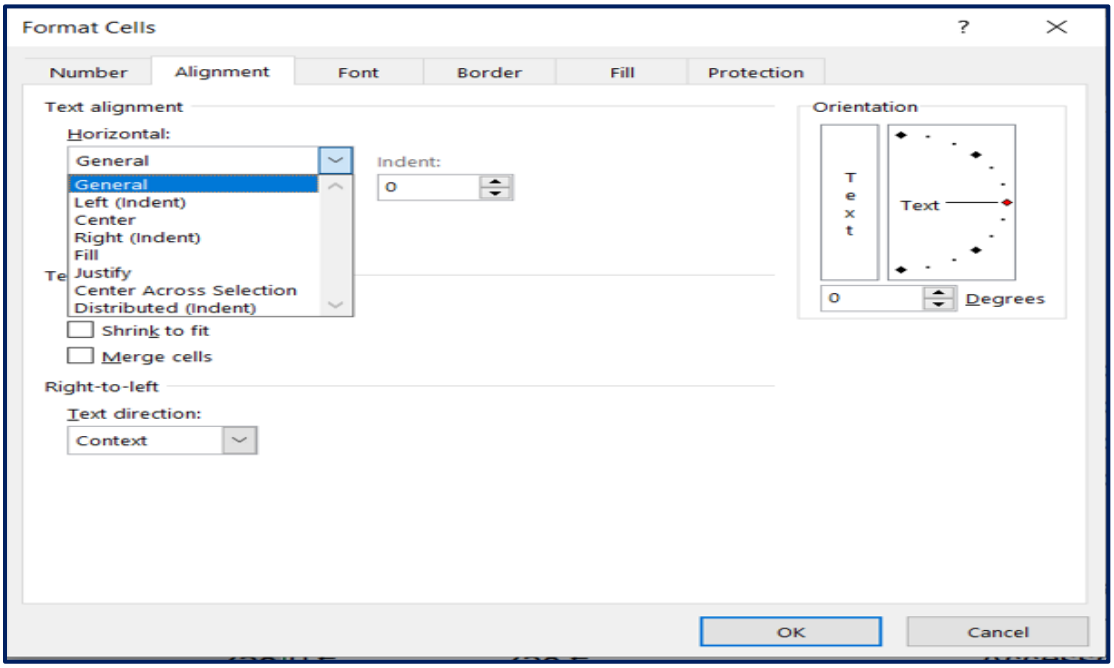
4. Discount => Number
This column will be converted from Currency to Number.

Steps:

1. convert from currency to Number from the home tap.



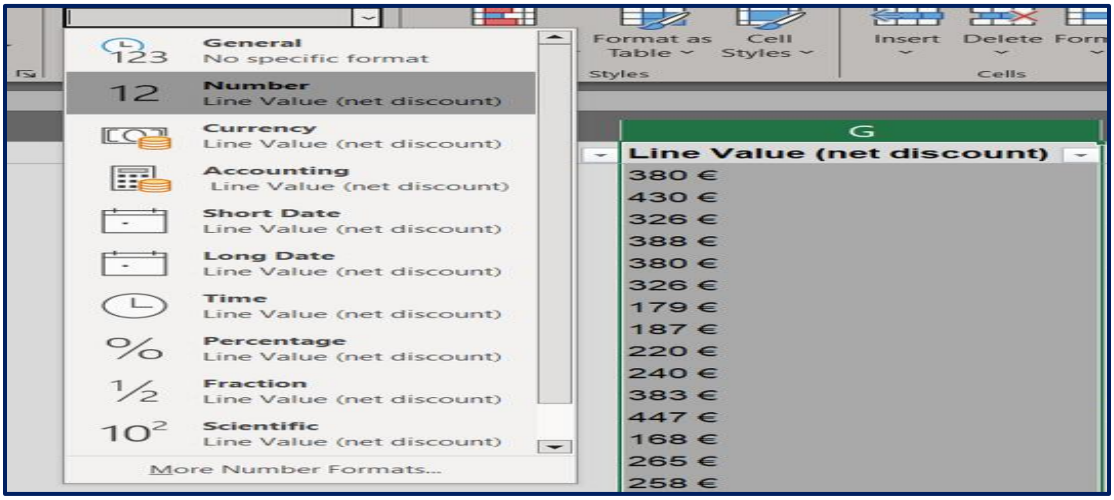
2. As the numbers are still aligned at left, to make it right aligned as the standard, we will make the alignment general instead of right.



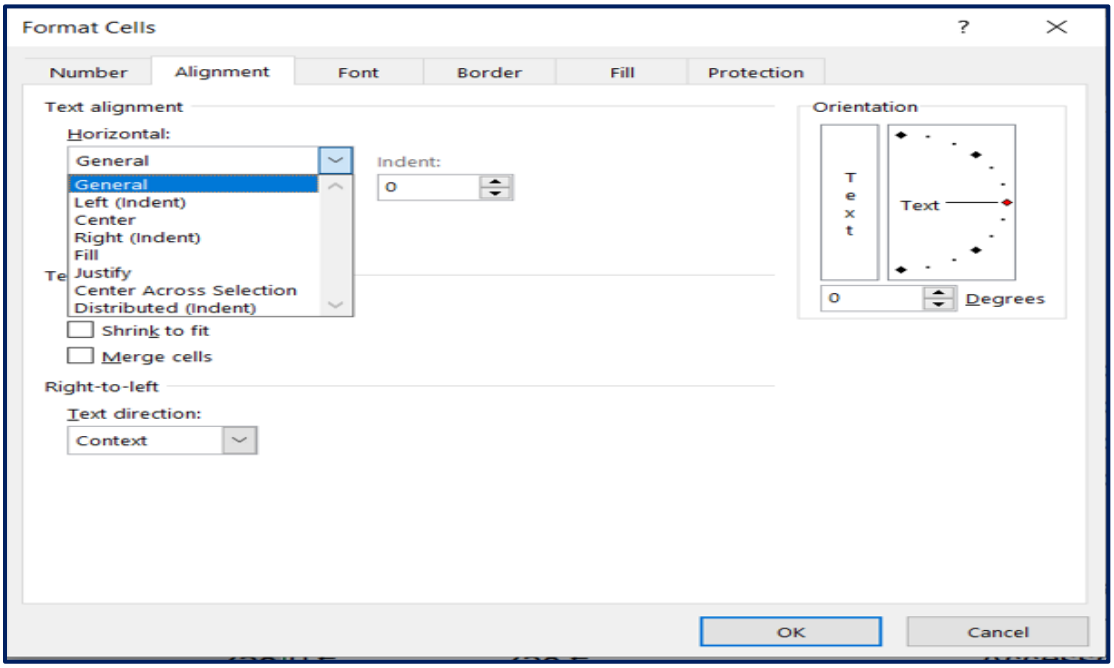
5. Line Value (net discount) => Number
   This column will be converted from Currency to Number.

   Steps:

   1. convert from currency to Number from the home tap.



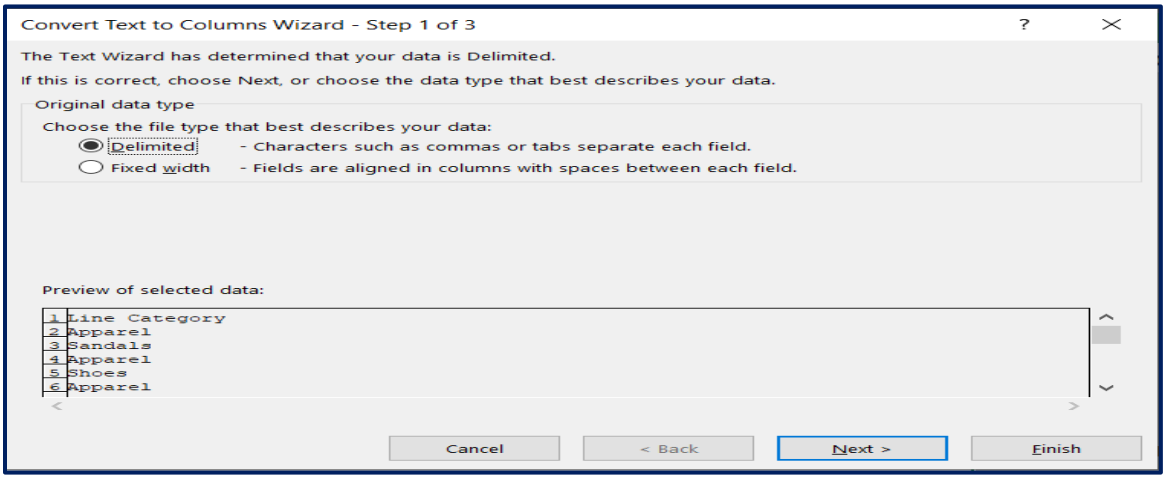
   2. As the numbers are still aligned at left, to make it right aligned as the standard, we will make the alignment general instead of right.

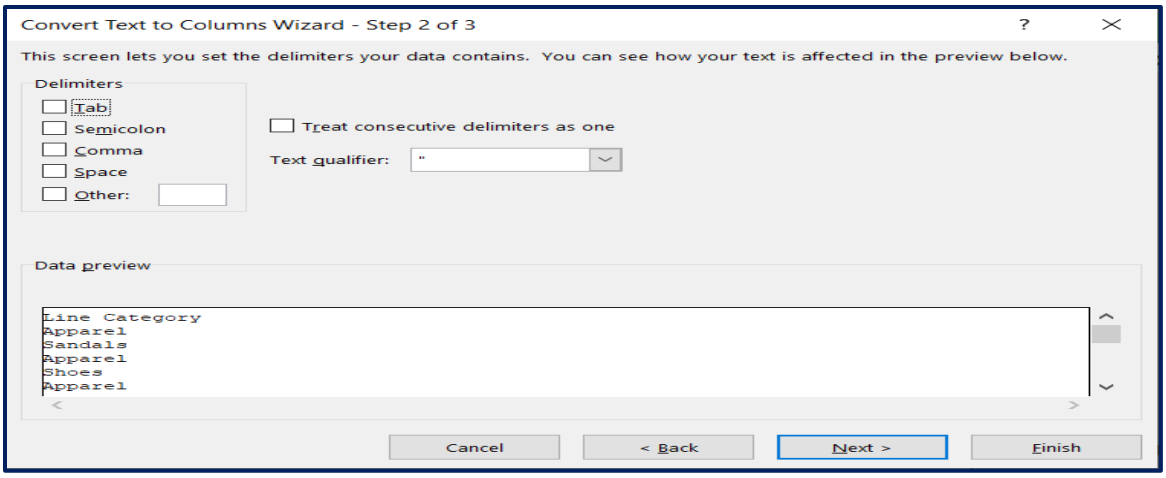


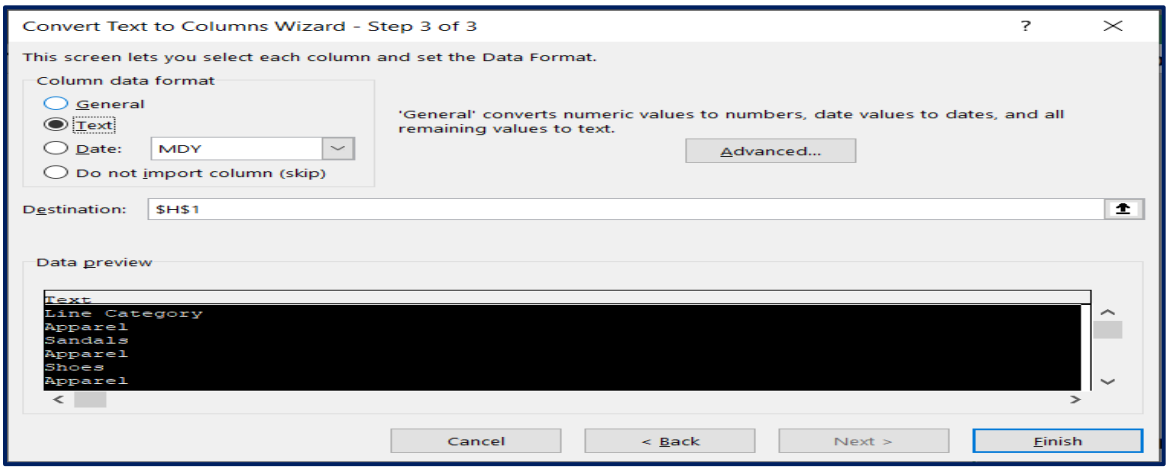
6. Line Category => Text

This column will be converted from General to Text using the **Text to Columns** feature.

NOTE: The Text to Columns feature was used rather than converting the type simply from the home tap, because it would require re-entering the data again.



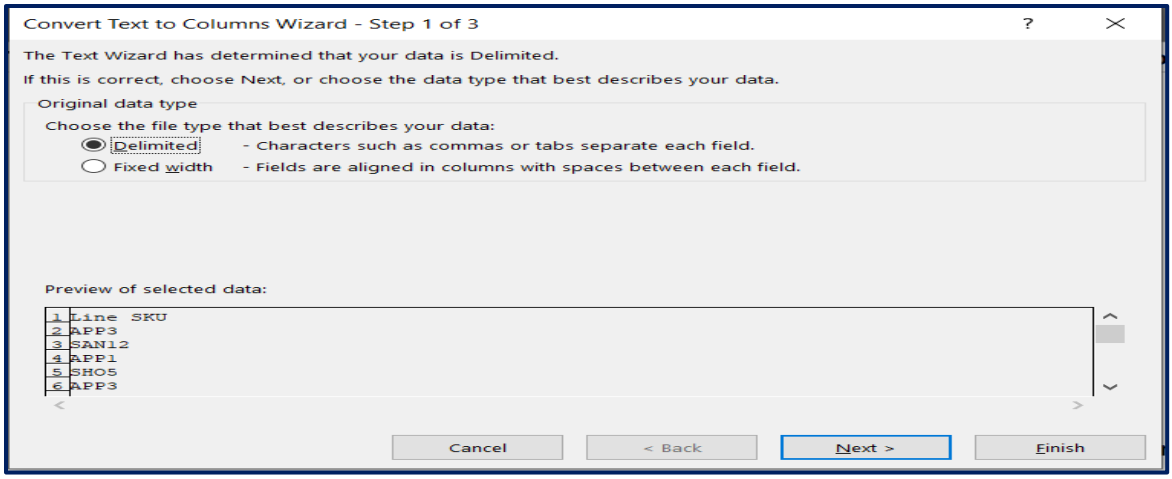


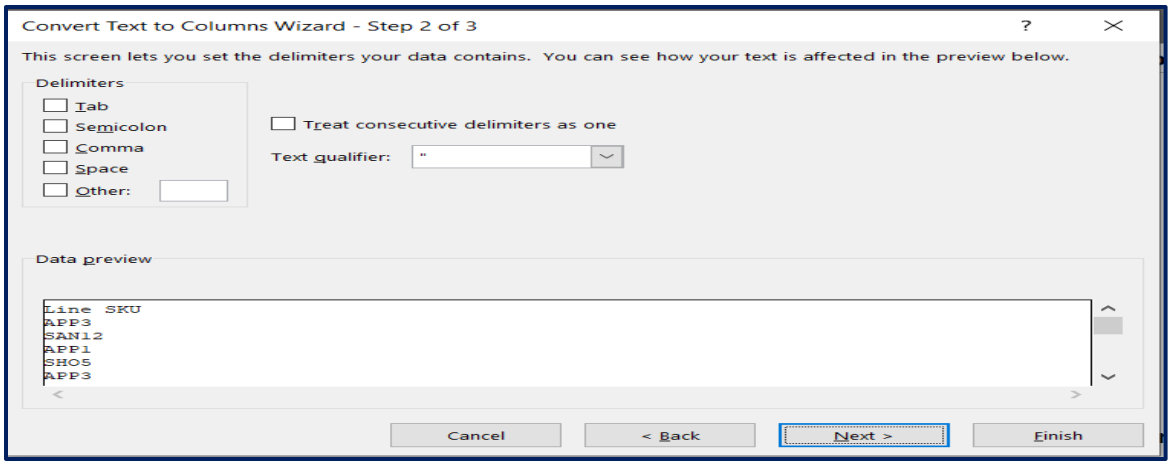


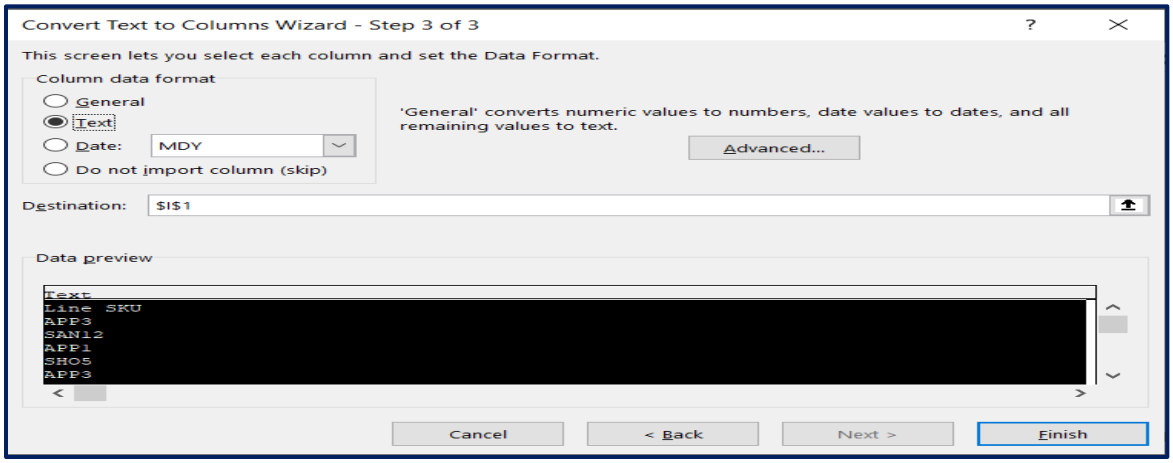
7. Line SKU => Text

This column will be converted from General to Text using the **Text to Columns** feature.

NOTE: The Text to Columns feature was used rather than converting the type simply from the home tap, because it would require re-entering the data again.

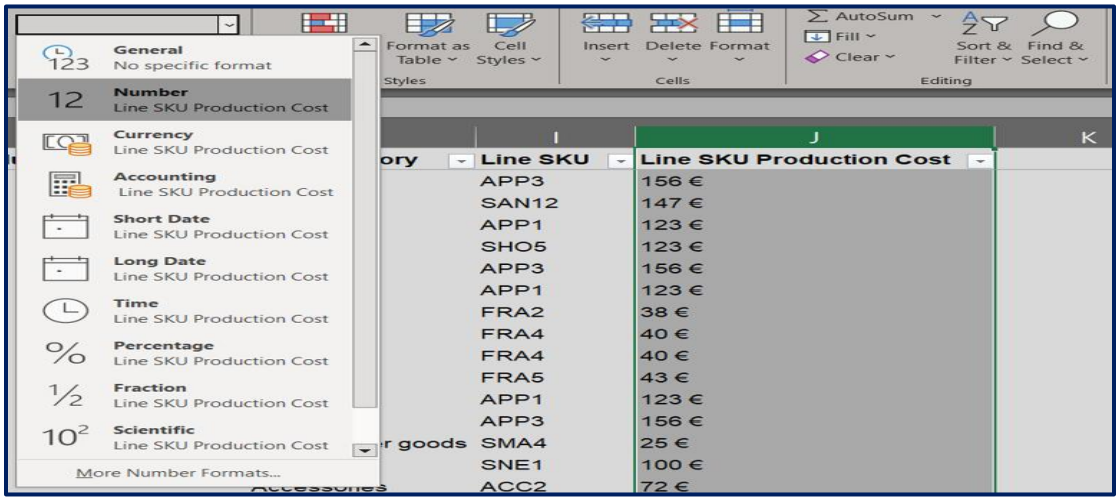




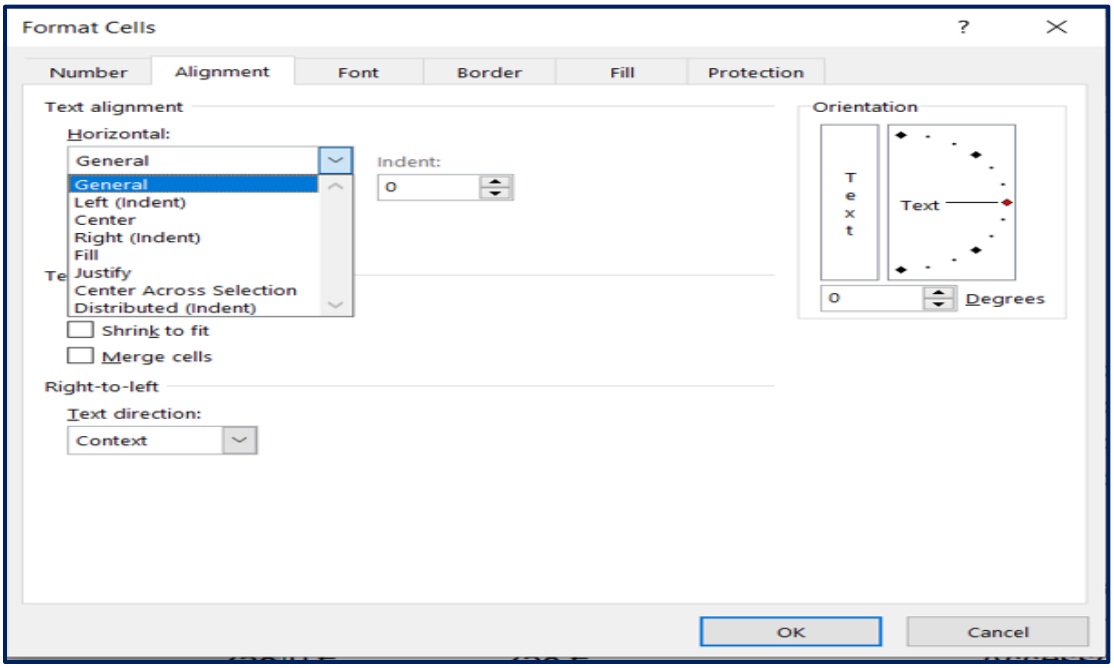


8. Line SKU Production Cost => Number
   This column will be converted from Currency to Number.

   Steps:
   1. convert from currency to Number from the home tap.



2. As the numbers are still aligned at left, to make it right aligned as the standard, we will make the alignment general instead of right.



Adjust and Format

In this process, we will add any needed columns for the analysis process.

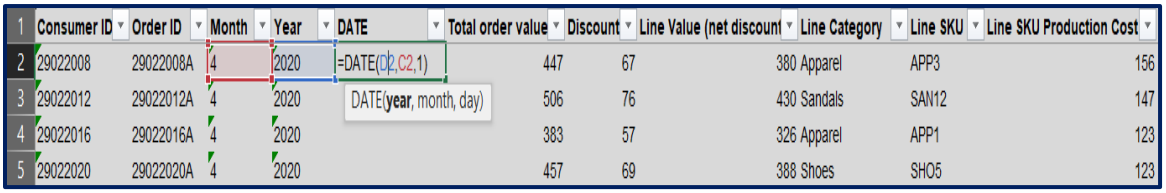
From the requirements, we need eight additional columns.

Field	Type	Description
Date	Date "m/d/yyyy"	Merging of the "Month" and "Year" columns.
Net Profit	Number	Equal the "Line Value (net discount)" Minus "Line SKU Production Cost".
Is New Consumer	Number	A Boolean variable equals '1' if this is the first order made by a customer, and equals '0' for any next order by the same customer.
Is Last Order	Number	A Boolean variable equals '1' if this is the last order made by a customer, and equals '0' otherwise.
Is Discount	Number	A Boolean variable equals '1' if the order is discounted, '0' otherwise.
Discount Category	General	A number representing the percentage of the order discount out of the total order value.
Number of Orders	Number	A Number representing the total number of orders purchased by a customer.
Order Number	Number	A Number representing the current order number purchased by a customer.

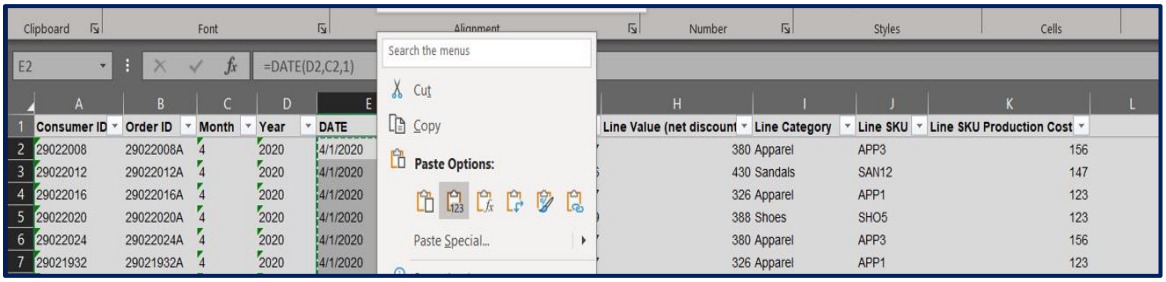
1. Date

Steps:

1. Create a new column called "Date" and use the **DATE** function.



2. Copy and paste as value.



3. Remove the “Month” and “Year” column, as they are no longer needed.

Consumer ID	Order ID	DATE	Total order value	Discount	Line Value (net discount	Line Category	Line SKU	Line SKU Production Cost
29022008	29022008A	4/1/2020	447	67	380	Apparel	APP3	156
29022012	29022012A	4/1/2020	506	76	430	Sandals	SAN12	147
29022016	29022016A	4/1/2020	383	57	326	Apparel	APP1	123
29022020	29022020A	4/1/2020	457	69	388	Shoes	SHO5	123

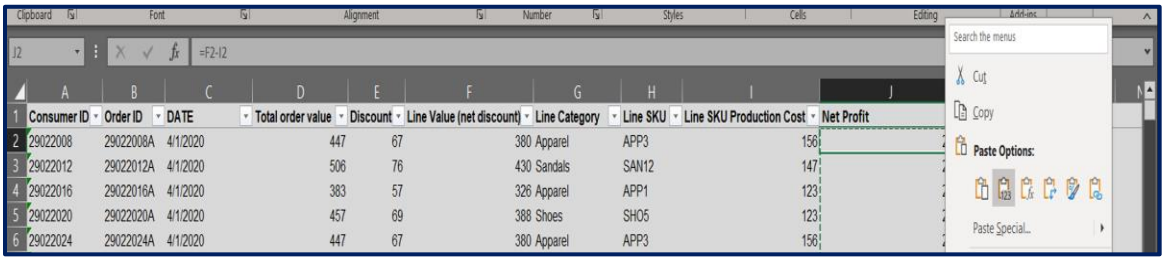
2. Net Profit

Steps:

1. Create a new column called “Net Profit” by applying the formula “Line Value (net discount)” minus “Line SKU Production Cost”.

Consumer ID	Order ID	DATE	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost	Net Profit
29022008	29022008A	4/1/2020	447	67	380	Apparel	APP3	156	=F2-I2
29022012	29022012A	4/1/2020	506	76	430	Sandals	SAN12	147	
29022016	29022016A	4/1/2020	383	57	326	Apparel	APP1	123	
29022020	29022020A	4/1/2020	457	69	388	Shoes	SHO5	123	

2. Copy and paste as value.

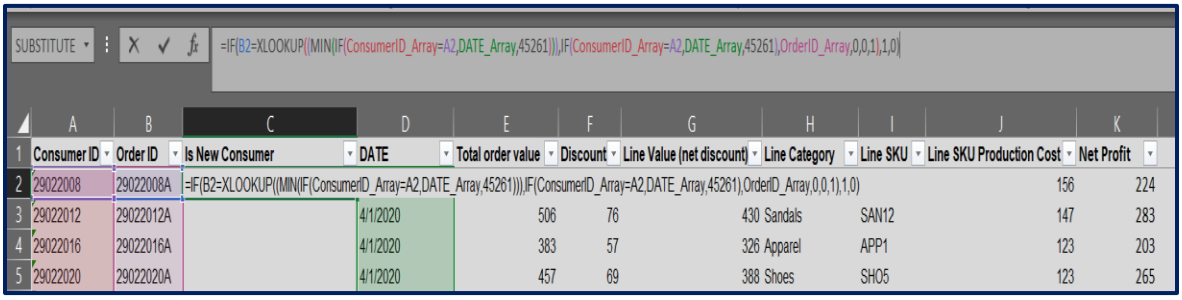


Consumer ID	Order ID	DATE	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost	Net Profit
29022008	29022008A	4/1/2020	447	67	380	Apparel	APP3	156	
29022012	29022012A	4/1/2020	506	76	430	Sandals	SAN12	147	
29022016	29022016A	4/1/2020	383	57	326	Apparel	APP1	123	
29022020	29022020A	4/1/2020	457	69	388	Shoes	SHO5	123	
29022024	29022024A	4/1/2020	447	67	380	Apparel	APP3	156	

3. Is New Consumer

Steps:

1. Create a new column called “Is New Consumer” and use **XLOOKUP**, **MIN**, and **IF** functions.



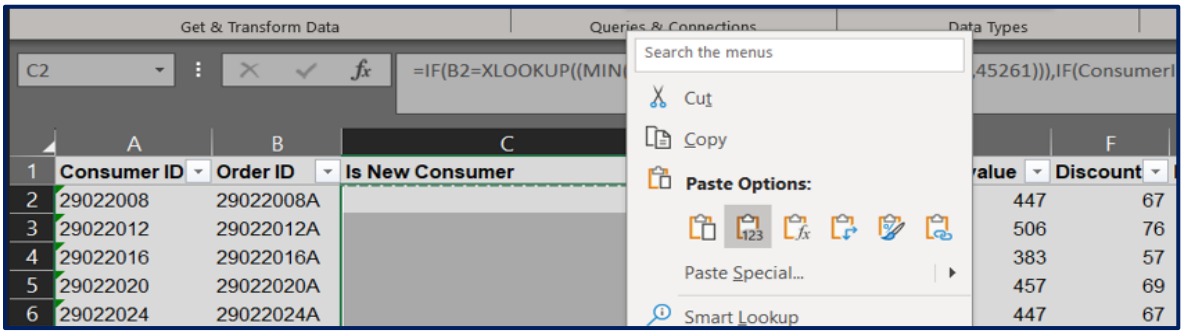
Consumer ID	Order ID	Is New Consumer	DATE	Total order value	Discount	Line Value (net discount)	Line Category	Line SKU	Line SKU Production Cost	Net Profit
29022008	29022008A	=IF(B2=XLOOKUP(MIN(IF(ConsumerID_Array=A2,DATE_Array,45261))),IF(ConsumerID_Array=A2,DATE_Array,45261),OrderID_Array,0,0,1),1,0)							156	224
29022012	29022012A		4/1/2020	506	76	430	Sandals	SAN12	147	283
29022016	29022016A		4/1/2020	383	57	326	Apparel	APP1	123	203
29022020	29022020A		4/1/2020	457	69	388	Shoes	SHO5	123	265

Explanation:

- **Intuition**: The function is designed to check if the current Order ID (B2) is the one associated with the earliest date for the current Consumer ID (A2). If it is, the function returns 1; otherwise, it returns 0. In case of a tie (i.e., multiple orders from the same Consumer ID have the same earliest date), the function picks the Order ID that appears first in the OrderID_Array.

- **XLOOKUP**(lookup_value, lookup_array, return_array, [if_not_found], [match_mode], [search_mode]):
    - lookup_value: This is the earliest date for the current Consumer ID (A2). It's found by taking the minimum date from the DATE_Array where the ConsumerID_Array equals A2. If the ConsumerID_Array doesn't equal A2, the function uses 45261 (which represents a date far in the future, assuming Excel's standard date system where 1 represents January 1, 1900).
    - lookup_array: This is the same array used to find the lookup_value.
    - return_array: This is the OrderID_Array from which the XLOOKUP function will return an Order ID.
    - [if_not_found]: If the lookup_value is not found in the lookup_array, the function returns 0.
    - [match_mode]: This is set to 0 to find an exact match.
    - [search_mode]: This is set to 1 to search from first to last.
- **IF**(logical_test, [value_if_true], [value_if_false]):
    - logical_test: This checks if the current Order ID (B2) is equal to the Order ID returned from the XLOOKUP function.
    - [value_if_true]: If the logical_test is true (i.e., the current Order ID is the one associated with the earliest date for the current Consumer ID), the function returns 1.
    - [value_if_false]: If the logical_test is false (i.e., the current Order ID is not the one associated with the earliest date for the current Consumer ID), the function returns 0.

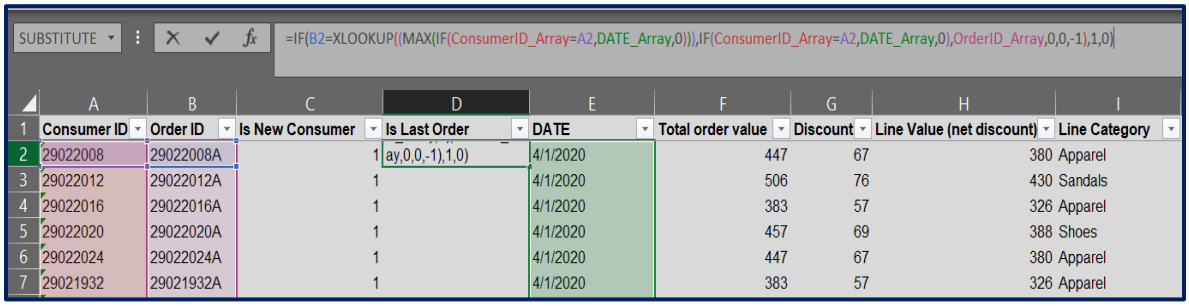
2. Copy and paste as value.



4. Is Last Order

Steps:

1. Create a new column called “Is Last Order” by using **XLOOKUP**, **MAX**, and **IF** functions.



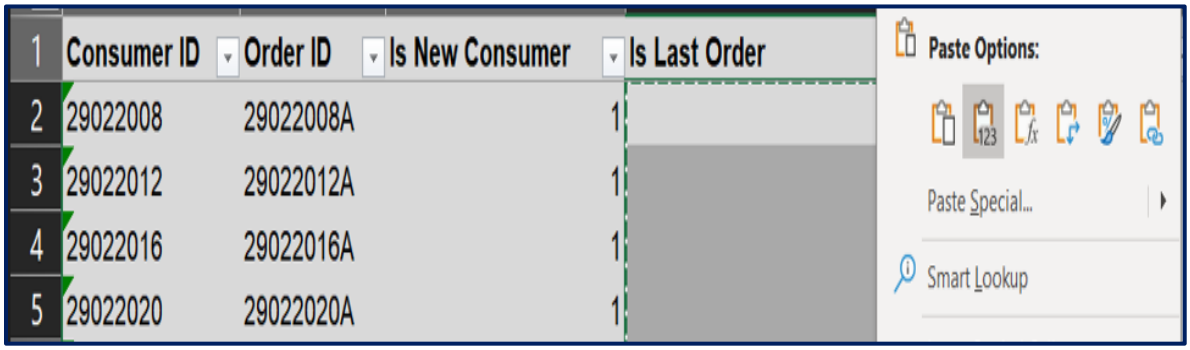
	A	B	C	D	E	F	G	H	I
1	Consumer ID	Order ID	Is New Consumer	Is Last Order	DATE	Total order value	Discount	Line Value (net discount)	Line Category
2	29022008	29022008A	1	ay,0,0,-1),1,0)	4/1/2020	447	67	380	Apparel
3	29022012	29022012A	1		4/1/2020	506	76	430	Sandals
4	29022016	29022016A	1		4/1/2020	383	57	326	Apparel
5	29022020	29022020A	1		4/1/2020	457	69	388	Shoes
6	29022024	29022024A	1		4/1/2020	447	67	380	Apparel
7	29021932	29021932A	1		4/1/2020	383	57	326	Apparel

Explanation:

- **Intuition**: The function is designed to check if the current Order ID (B2) is the one associated with the latest date for the current Consumer ID (A2). If it is, the function returns 1; otherwise, it returns 0. In case of a tie (i.e., multiple orders from the same Consumer ID have the same latest date), the function picks the Order ID that appears last in the OrderID_Array.

- **XLOOKUP**(lookup_value, lookup_array, return_array, [if_not_found], [match_mode], [search_mode]):
  - lookup_value: This is the latest date for the current Consumer ID (A2). It’s found by taking the maximum date from the DATE_Array where the ConsumerID_Array equals A2. If the ConsumerID_Array doesn’t equal A2, the function uses 0.
  - lookup_array: This is the same array used to find the lookup_value.
  - return_array: This is the OrderID_Array from which the XLOOKUP function will return an Order ID.
  - [if_not_found]: If the lookup_value is not found in the lookup_array, the function returns 0.
  - [match_mode]: This is set to 0 to find an exact match.
  - [search_mode]: This is set to -1 to search from last to first.
- **IF**(logical_test, [value_if_true], [value_if_false]):
  - logical_test: This checks if the current Order ID (B2) is equal to the Order ID returned from the XLOOKUP function.
  - [value_if_true]: If the logical_test is true (i.e., the current Order ID is the one associated with the latest date for the current Consumer ID), the function returns 1.
  - [value_if_false]: If the logical_test is false (i.e., the current Order ID is not the one associated with the latest date for the current Consumer ID), the function returns 0.

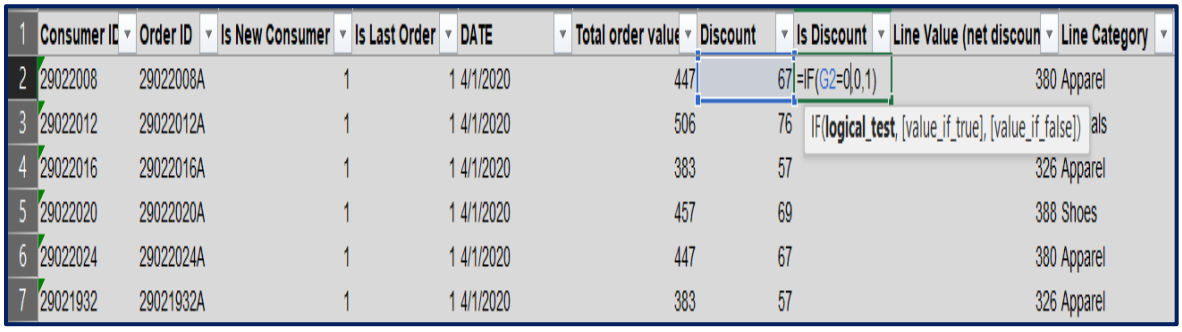
2. Copy and paste as value.



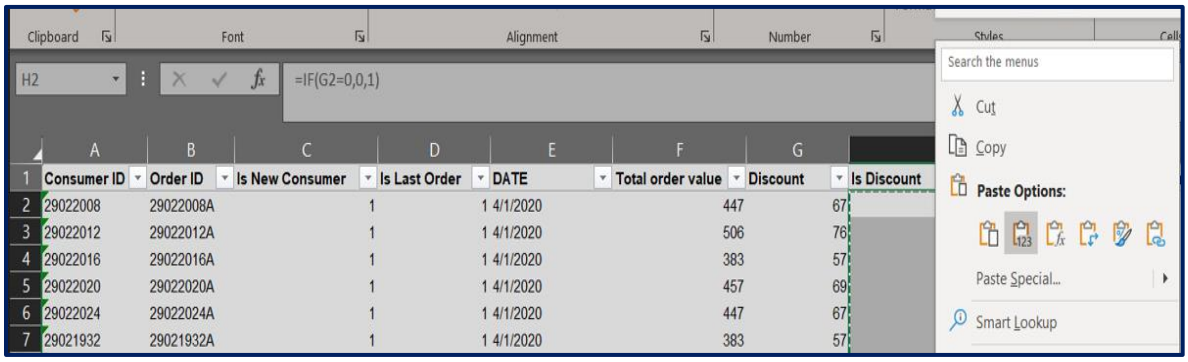
5. Is Discount

Steps:

1. Create a new column called “Is Discount” by using the **IF** function.



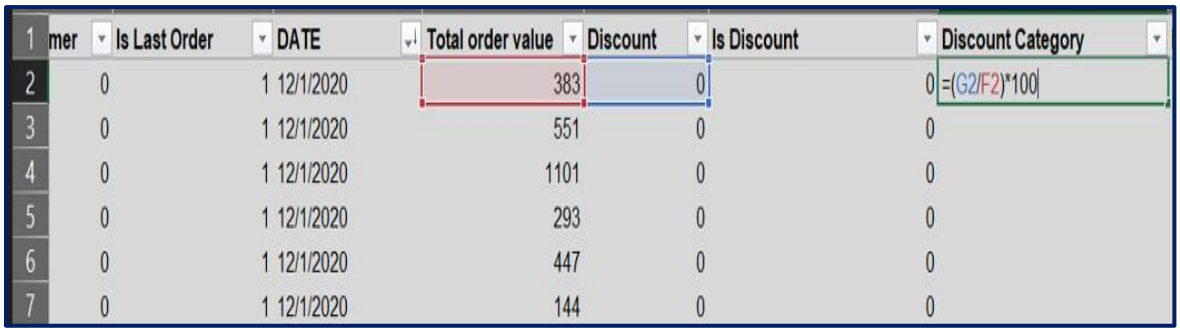
2. Copy and paste as value.



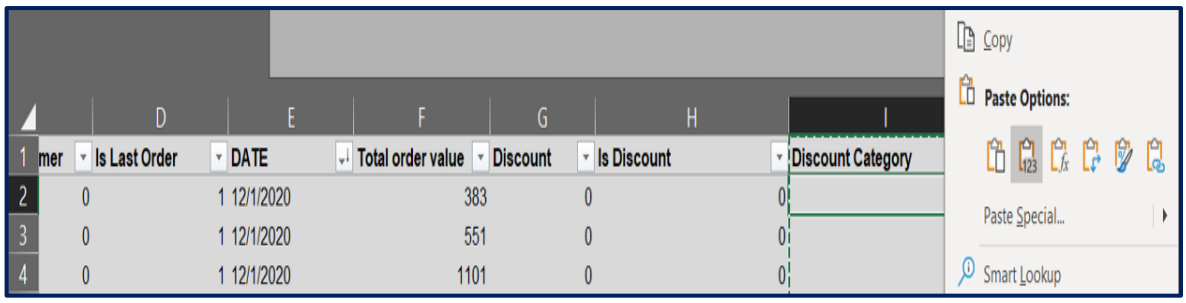
6. Discount Category

Steps:

1. Create a new column called “Discount Category” by applying the formula (Discount/Total order value)*100.



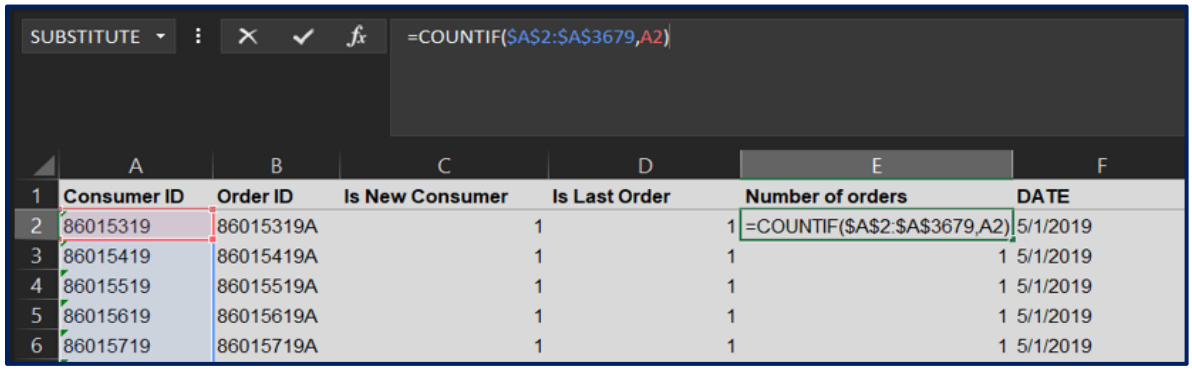
2. Copy and paste as value.



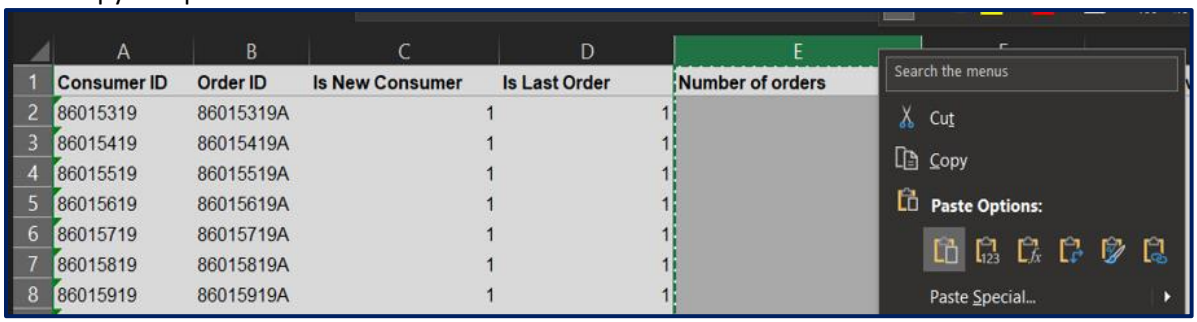
7. Number of Orders

Steps:

1. Create a new column called “Number of Orders” by using **COUNTIF** functions.



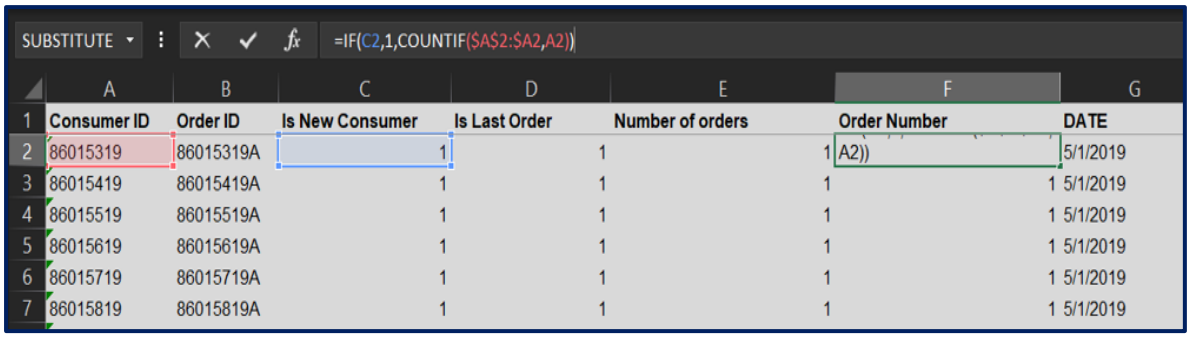
2. Copy and paste as value.



7. Order Number

Steps:

1. Create a new column called “Number of Orders” by using **COUNTIF** and **IF** functions.



2. Copy and paste as value.