VSIT

Business Intelligence
Practical #2

Name	Sanjeev Gupta	**Roll Number**	21302B0023
Class	TYBScIT	**Division**	C
Subject/Course:	Business intelligence		
Topic	Perform Data Wrangling (ETL)		

What is ETL?

ETL, which stands for extract, transform and load, is a data integration process that combines data from multiple data sources into a single, consistent data store that is loaded into a data warehouse or other target system.

As the databases grew in popularity in the 1970s, ETL was introduced as a process for integrating and loading data for computation and analysis, eventually becoming the primary method to process data for data warehousing projects.

ETL provides the foundation for data analytics and machine learning workstreams. Through a series of business rules, ETL cleanses and organizes data in a way which addresses specific business intelligence needs, like monthly reporting, but it can also tackle more advanced analytics, which can improve back-end processes or end user experiences. ETL is often used by an organization to:

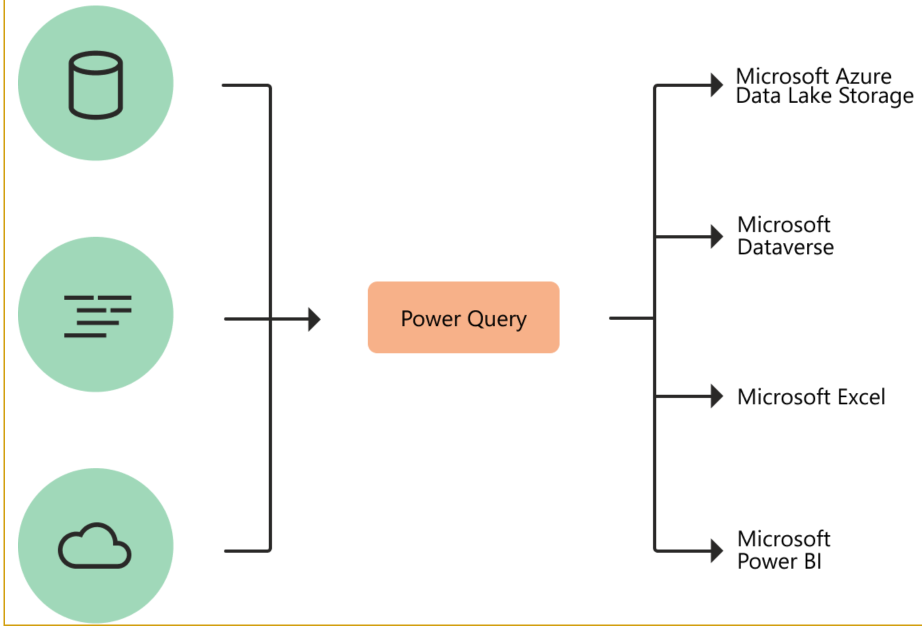
- Extract data from legacy systems
- Cleanse the data to improve data quality and establish consistency
- Load data into a target database

WHY ETL USED IN Data Science PROJECTS:

ETL is essential for organizations that wish to integrate data from multiple sources, such as databases, APIs, and flat files, into a single repository or data warehouse, and transform it into a format that is suitable for analysis or business needs. Overall, ETL is necessary for managing data at scale, ensuring data quality and accuracy, and enabling businesses to make informed decisions based on the data.

What is Power Query?

Power Query is a data transformation and data preparation engine. Power Query comes with a graphical interface for getting data from sources and a Power Query Editor for applying transformations. Because the engine is available in many products and services, the destination where the data will be stored depends on where Power Query was used. Using Power Query, you can perform the extract, transform, and load (ETL) processing of data.



What are the components of Power Query Editor?

The Power Query Editor is a versatile tool within Microsoft Excel that allows users to clean, transform, and combine data from various sources. It is an essential part of the Power BI suite and is used to create and manage data for data analysis and visualization. The components of the Power Query Editor can be broadly categorized into four main sections:

Data Source Navigator: This is the starting point for importing data into the Power Query Editor. It allows users to browse through the available data sources and import the desired data into the editor. Data sources can include Excel files, text files, databases, web pages, and more.

Query Settings: This section provides options to configure the imported data, such as renaming columns, changing data types, and filtering rows. Users can also perform various transformations on the data, such as splitting columns, merging columns, and removing duplicates.

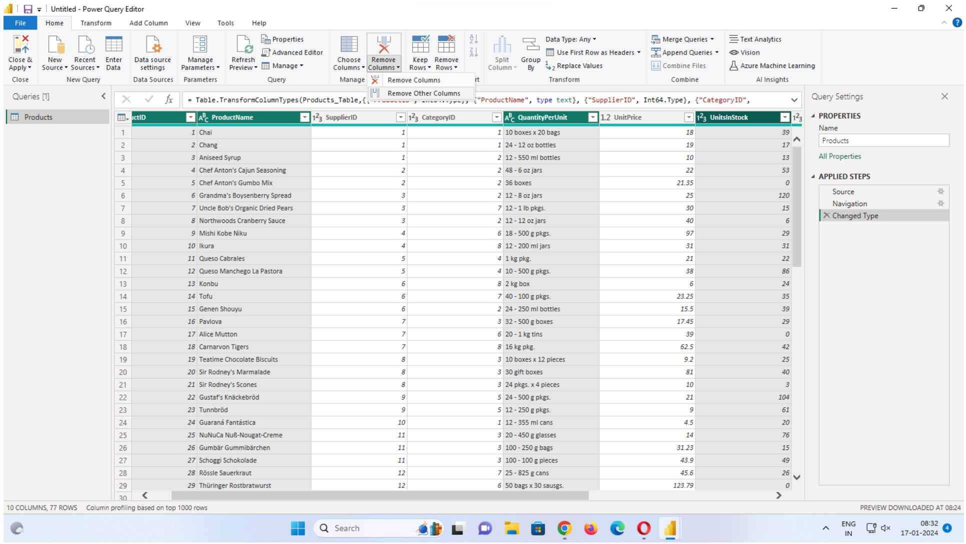
Transformation and Actions: The Power Query Editor provides a range of built-in functions and transformations to modify the data. Users can also create custom functions using the M language, which is a functional programming language specifically designed for data manipulation. Some common transformations include sorting, grouping, and calculating aggregates.

Advanced Editor: This section allows users to view and edit the underlying M code that represents the data transformation process. Advanced users can modify the M code directly or use it as a reference for understanding the data manipulation process.

In summary, the Power Query Editor is a powerful tool for data manipulation and transformation, consisting of the Data Source Navigator, Query Settings, Transformation and Actions, and the Advanced Editor. These components work together to enable users to clean, transform, and combine data from various sources for analysis and visualization.

Write the steps to perform ETL in Power BI?

1. Remove other columns to only display columns of interest.



```
= Table.TransformColumnTypes(Products_Table,{{"ProductID", Int64.Type}, {"ProductName", type text}, {"SupplierID", Int64.Type}, {"CategoryID",
```

Name: Products

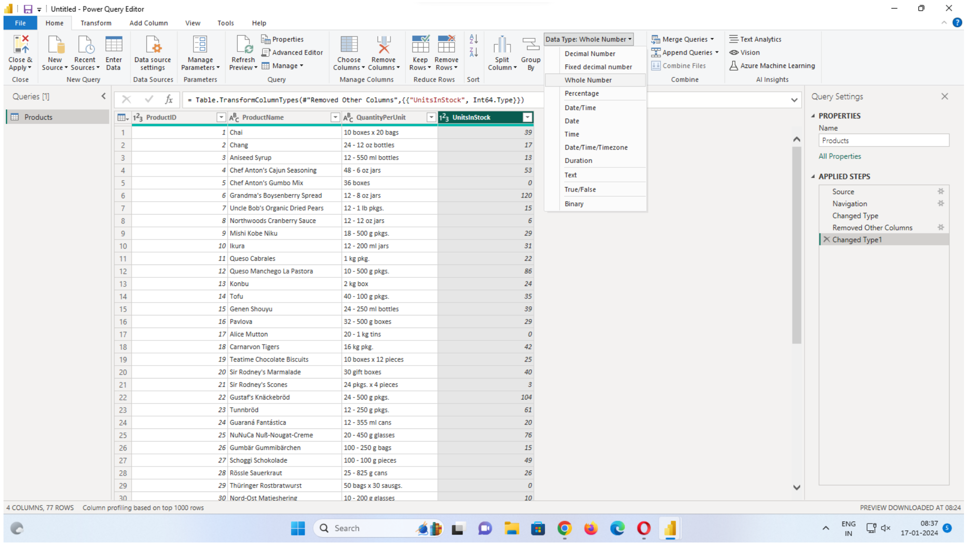
APPLIED STEPS: Source, Navigation, Changed Type

	ctID	ProductName	SupplierID	CategoryID	QuantityPerUnit	UnitPrice	UnitsInStock
1	1	Chai	1	1	10 boxes x 20 bags	18	39
2	2	Chang	1	1	24 - 12 oz bottles	19	17
3	3	Aniseed Syrup	1	2	12 - 550 ml bottles	10	13
4	4	Chef Anton's Cajun Seasoning	2	2	48 - 6 oz jars	22	53
5	5	Chef Anton's Gumbo Mix	2	2	36 boxes	21.35	0
6	6	Grandma's Boysenberry Spread	3	2	12 - 8 oz jars	25	120
7	7	Uncle Bob's Organic Dried Pears	3	7	12 - 1 lb pkgs.	30	15
8	8	Northwoods Cranberry Sauce	3	2	12 - 12 oz jars	40	6
9	9	Mishi Kobe Niku	4	6	18 - 500 g pkgs.	97	29
10	10	Ikura	4	8	12 - 200 ml jars	31	31
11	11	Queso Cabrales	5	4	1 kg pkg.	21	22
12	12	Queso Manchego La Pastora	5	4	10 - 500 g pkgs.	38	86
13	13	Konbu	6	8	2 kg box	6	24
14	14	Tofu	6	7	40 - 100 g pkgs.	23.25	35
15	15	Genen Shouyu	6	2	24 - 250 ml bottles	15.5	39
16	16	Pavlova	7	3	32 - 500 g boxes	17.45	29
17	17	Alice Mutton	7	6	20 - 1 kg tins	39	0
18	18	Carnarvon Tigers	7	8	16 kg pkg.	62.5	42
19	19	Teatime Chocolate Biscuits	8	3	10 boxes x 12 pieces	9.2	25
20	20	Sir Rodney's Marmalade	8	3	30 gift boxes	81	40
21	21	Sir Rodney's Scones	8	3	24 pkgs. x 4 pieces	10	3
22	22	Gustaf's Knäckebröd	9	5	24 - 500 g pkgs.	21	104
23	23	Tunnbröd	9	5	12 - 250 g pkgs.	9	61
24	24	Guaraná Fantástica	10	1	12 - 355 ml cans	4.5	20
25	25	NuNuCa Nuß-Nougat-Creme	11	3	20 - 450 g glasses	14	76
26	26	Gumbär Gummibärchen	11	3	100 - 250 g bags	31.23	15
27	27	Schoggi Schokolade	11	3	100 - 100 g pieces	43.9	49
28	28	Rössle Sauerkraut	12	7	25 - 825 g cans	45.6	26
29	29	Thüringer Rostbratwurst	12	6	50 bags x 30 sausgs.	123.79	0
30							

10 COLUMNS, 77 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 08:24

2. Change the data type in UnitInStock column.



```
= Table.TransformColumnTypes(#"Removed Other Columns",{{"UnitsInStock", Int64.Type}})
```

Name: Products

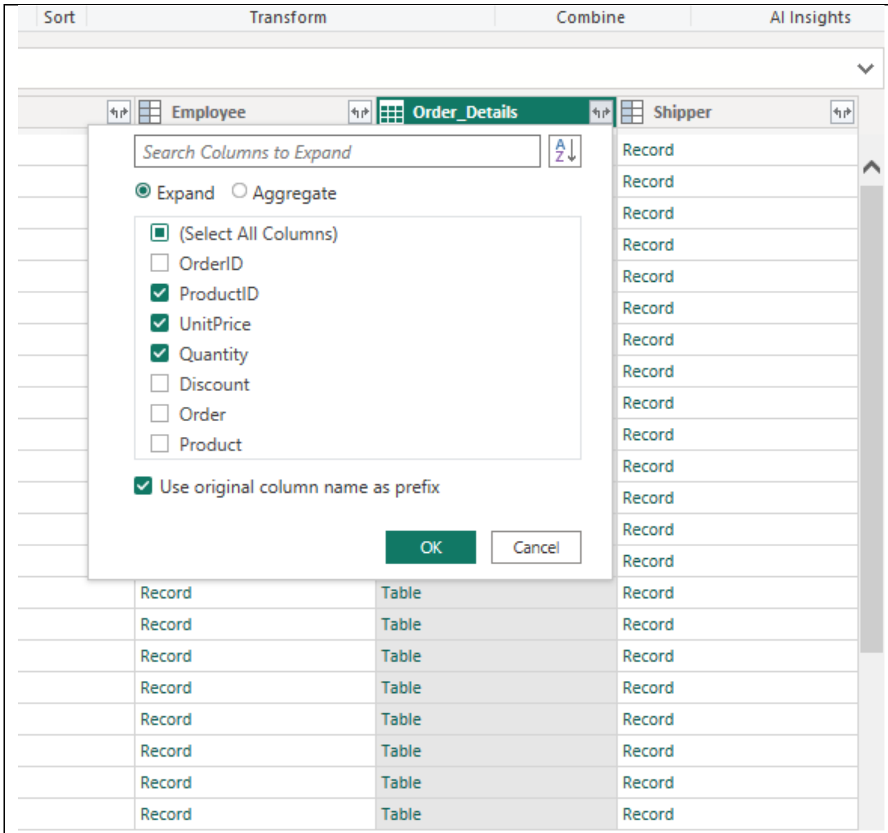
APPLIED STEPS: Source, Navigation, Changed Type, Removed Other Columns, Changed Type1

	ProductID	ProductName	QuantityPerUnit	UnitsInStock
1	1	Chai	10 boxes x 20 bags	39
2	2	Chang	24 - 12 oz bottles	17
3	3	Aniseed Syrup	12 - 550 ml bottles	13
4	4	Chef Anton's Cajun Seasoning	48 - 6 oz jars	53
5	5	Chef Anton's Gumbo Mix	36 boxes	0
6	6	Grandma's Boysenberry Spread	12 - 8 oz jars	120
7	7	Uncle Bob's Organic Dried Pears	12 - 1 lb pkgs.	15
8	8	Northwoods Cranberry Sauce	12 - 12 oz jars	6
9	9	Mishi Kobe Niku	18 - 500 g pkgs.	29
10	10	Ikura	12 - 200 ml jars	31
11	11	Queso Cabrales	1 kg pkg.	22
12	12	Queso Manchego La Pastora	10 - 500 g pkgs.	86
13	13	Konbu	2 kg box	24
14	14	Tofu	40 - 100 g pkgs.	35
15	15	Genen Shouyu	24 - 250 ml bottles	39
16	16	Pavlova	32 - 500 g boxes	29
17	17	Alice Mutton	20 - 1 kg tins	0
18	18	Carnarvon Tigers	16 kg pkg.	42
19	19	Teatime Chocolate Biscuits	10 boxes x 12 pieces	25
20	20	Sir Rodney's Marmalade	30 gift boxes	40
21	21	Sir Rodney's Scones	24 pkgs. x 4 pieces	3
22	22	Gustaf's Knäckebröd	24 - 500 g pkgs.	104
23	23	Tunnbröd	12 - 250 g pkgs.	61
24	24	Guaraná Fantástica	12 - 355 ml cans	20
25	25	NuNuCa Nuß-Nougat-Creme	20 - 450 g glasses	76
26	26	Gumbär Gummibärchen	100 - 250 g bags	15
27	27	Schoggi Schokolade	100 - 100 g pieces	49
28	28	Rössle Sauerkraut	25 - 825 g cans	26
29	29	Thüringer Rostbratwurst	50 bags x 30 sausgs.	0
30	30	Nord-Ost Matjeshering	10 - 200 g glasses	10

4 COLUMNS, 77 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 08:24

3. Expand the Order_Details table.



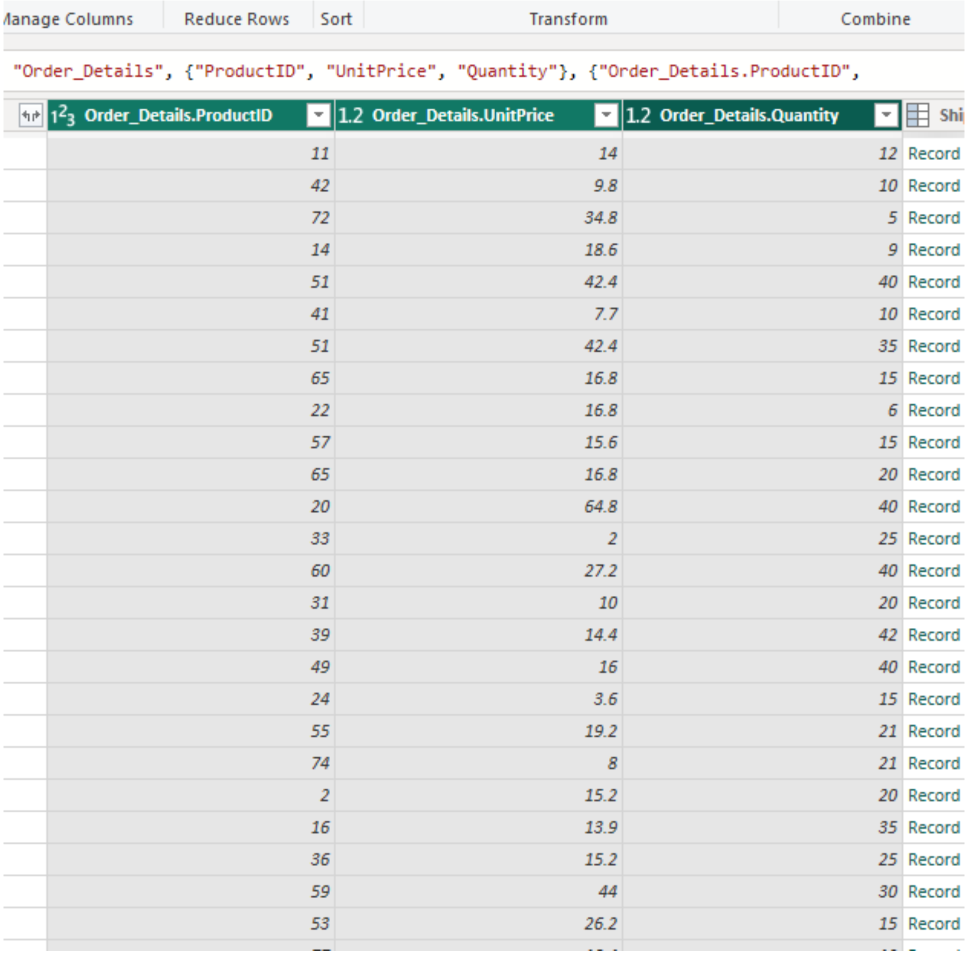
Search Columns to Expand

● Expand ○ Aggregate

- ▣ (Select All Columns)
- □ OrderID
- ☑ ProductID
- ☑ UnitPrice
- ☑ Quantity
- □ Discount
- □ Order
- □ Product

☑ Use original column name as prefix

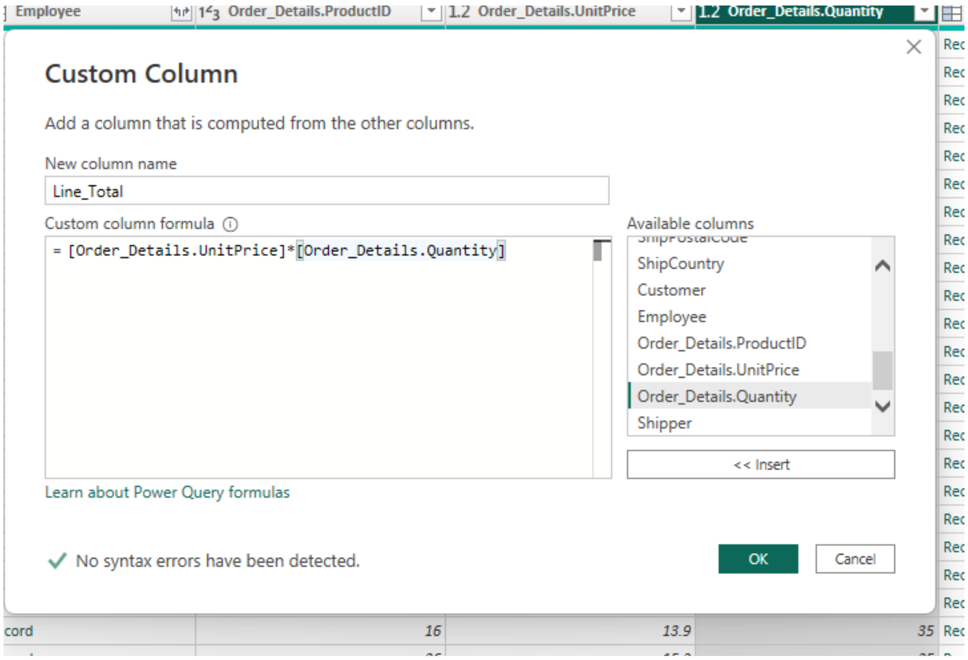
Employee	Order_Details	Shipper
		Record
		Record
		Record
		Record
		Record
		Record
		Record
		Record
		Record
		Record
		Record
		Record
		Record
		Record
Record	Table	Record
Record	Table	Record
Record	Table	Record
Record	Table	Record
Record	Table	Record
Record	Table	Record
Record	Table	Record
Record	Table	Record



```
"Order_Details", {"ProductID", "UnitPrice", "Quantity"}, {"Order_Details.ProductID",
```

1²3 Order_Details.ProductID	1.2 Order_Details.UnitPrice	1.2 Order_Details.Quantity	Shi
11	14	12	Record
42	9.8	10	Record
72	34.8	5	Record
14	18.6	9	Record
51	42.4	40	Record
41	7.7	10	Record
51	42.4	35	Record
65	16.8	15	Record
22	16.8	6	Record
57	15.6	15	Record
65	16.8	20	Record
20	64.8	40	Record
33	2	25	Record
60	27.2	40	Record
31	10	20	Record
39	14.4	42	Record
49	16	40	Record
24	3.6	15	Record
55	19.2	21	Record
74	8	21	Record
2	15.2	20	Record
16	13.9	35	Record
36	15.2	25	Record
59	44	30	Record
53	26.2	15	Record

4. Calculate the line total for each Order_Details row.


Employee
1²3 Order_Details.ProductID
1.2 Order_Details.UnitPrice
1.2 Order_Details.Quantity

Custom Column

Add a column that is computed from the other columns.

New column name

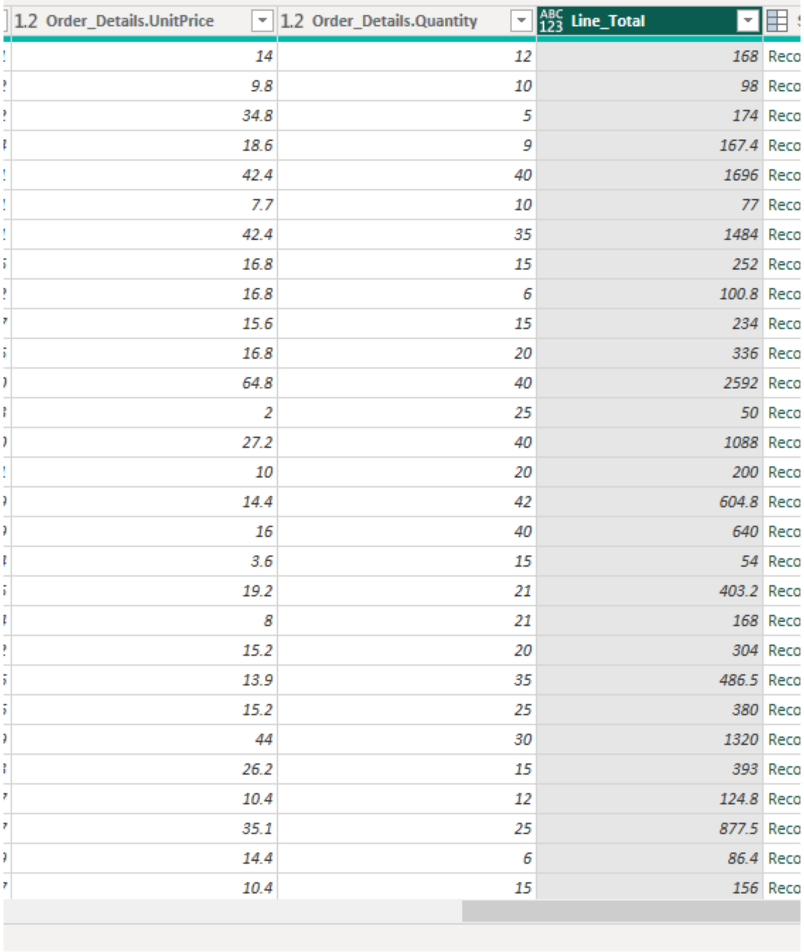
Line_Total

Custom column formula

```
= [Order_Details.UnitPrice]*[Order_Details.Quantity]
```

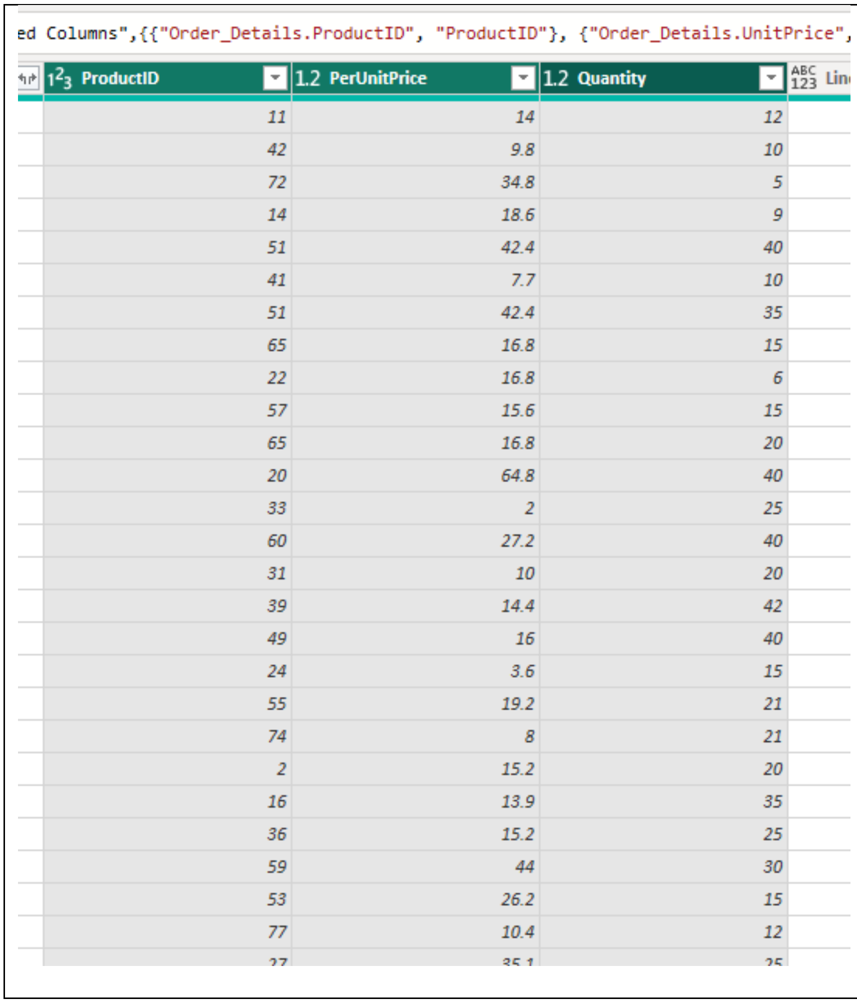
Available columns

✓ No syntax errors have been detected.



1.2 Order_Details.UnitPrice	1.2 Order_Details.Quantity	ABC 123 Line_Total	
14	12	168	Reco
9.8	10	98	Reco
34.8	5	174	Reco
18.6	9	167.4	Reco
42.4	40	1696	Reco
7.7	10	77	Reco
42.4	35	1484	Reco
16.8	15	252	Reco
16.8	6	100.8	Reco
15.6	15	234	Reco
16.8	20	336	Reco
64.8	40	2592	Reco
2	25	50	Reco
27.2	40	1088	Reco
10	20	200	Reco
14.4	42	604.8	Reco
16	40	640	Reco
3.6	15	54	Reco
19.2	21	403.2	Reco
8	21	168	Reco
15.2	20	304	Reco
13.9	35	486.5	Reco
15.2	25	380	Reco
44	30	1320	Reco
26.2	15	393	Reco
10.4	12	124.8	Reco
35.1	25	877.5	Reco
14.4	6	86.4	Reco
10.4	15	156	Reco

5. Rename and Reorder columns in the query



```
ed Columns",{{"Order_Details.ProductID", "ProductID"}, {"Order_Details.UnitPrice",
```

ProductID	PerUnitPrice	Quantity	Lin
11	14	12	
42	9.8	10	
72	34.8	5	
14	18.6	9	
51	42.4	40	
41	7.7	10	
51	42.4	35	
65	16.8	15	
22	16.8	6	
57	15.6	15	
65	16.8	20	
20	64.8	40	
33	2	25	
60	27.2	40	
31	10	20	
39	14.4	42	
49	16	40	
24	3.6	15	
55	19.2	21	
74	8	21	
2	15.2	20	
16	13.9	35	
36	15.2	25	
59	44	30	
53	26.2	15	
77	10.4	12	
27	35.1	25	