

Exploring the Power Query

In this case, we will explore Power Query – learn how to transform data before final loading it for analysis. This tutorial helped me understand how to obtain and clean up data in Microsoft Excel and Microsoft Power BI using Microsoft Power Query. The simplest approach to connect to, take data from, transform it, and load it is with Power Query.

power query - Power Query Editor

File Home Transform Add Column View

Close & Load Refresh Preview Advanced Editor Manage Query

Choose Columns Remove Columns Keep Rows Remove Rows Sort Split Column Group By Data Type: Text Use First Row as Headers Merge Queries Append Queries Combine Files Manage Parameters Data source settings New Source Recent Sources Enter Data New Query

Queries [5]

Transform File from... Helper Queries [3] Parameter1 (Sa... Sample File Transform File Transform Sampl... Other Queries [1] power query

Table.TransformColumnTypes(#"Expanded Table Column1",{{"Source.Name", type text}, {"Order ID", type any},

	Source.Name	Order ID	Customer ID	Cookies Shipped	Revenue
1	2017 Order Data.xlsx	413383	152689 - YT Restaurants	936	4680
2	2017 Order Data.xlsx	751142	985245 - Quick Bite Convenience Stores	987	4935
3	2017 Order Data.xlsx	876056	985245 - Quick Bite Convenience Stores	270	1350
4	2017 Order Data.xlsx	548247	985245 - Quick Bite Convenience Stores	925	4625
5	2017 Order Data.xlsx	626350	152689 - YT Restaurants	974	4870
6	2017 Order Data.xlsx	XXXXXX	XXXXXX - Promo	715	3575
7	2017 Order Data.xlsx	200181	325698 - Cascade Grovers	454	2270
8	2017 Order Data.xlsx	711368	152689 - YT Restaurants	484	2420
9	2017 Order Data.xlsx	958419	325698 - Cascade Grovers	269	1345
10	2017 Order Data.xlsx	808349	985245 - Quick Bite Convenience Stores	396	1980
11	2017 Order Data.xlsx	392582	985245 - Quick Bite Convenience Stores	121	605
12	2017 Order Data.xlsx	415147	152689 - YT Restaurants	373	1865
13	2017 Order Data.xlsx	656789	325698 - Cascade Grovers	118	590
14	2017 Order Data.xlsx	247229	325698 - Cascade Grovers	137	685
15	2017 Order Data.xlsx	532184	985245 - Quick Bite Convenience Stores	574	2870
16	2017 Order Data.xlsx	412714	325698 - Cascade Grovers	541	2705
17	2017 Order Data.xlsx	771432	325698 - Cascade Grovers	932	4660
18	2017 Order Data.xlsx	363695	785432 - Park & Eat LLC	172	860
19	2017 Order Data.xlsx	395821	452584 - Acme Grocery Stores	647	3235
20	2017 Order Data.xlsx	234191	325698 - Cascade Grovers	660	3300
21	2017 Order Data.xlsx	584039	325698 - Cascade Grovers	907	4535
22	2017 Order Data.xlsx	463601	325698 - Cascade Grovers	429	2145
23	2017 Order Data.xlsx	826310	325698 - Cascade Grovers	645	3225

9 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 9:46 AM

Query Settings

PROPERTIES

Name

power query

APPLIED STEPS

Source

Filtered Hidden Files1

Invoke Custom Function1

Renamed Columns1

Removed Other Columns1

Expanded Table Column1

Changed Type

Since I don't want source column, I simply removed this column.

power query - Power Query Editor

File Home Transform Add Column View

Close & Load Refresh Preview Advanced Editor Manage Query

Choose Columns Remove Columns Manage Columns

Keep Rows Remove Rows Reduce Rows

Sort Split Column Group By

Data Type: Any Use First Row as Headers Replace Values

Merge Queries Append Queries Combine Files

Manage Parameters Combine Parameters

Data source settings Data Sources

New Source Recent Sources Enter Data New Query

Queries [5]

- Transform File from...
- Helper Queries [3]
 - Parameter1 (Sa...
 - Sample File
 - Transform File
 - Transform Sampl...
- Other Queries [1]
 - power query

fx = Table.RemoveColumns(#"Changed Type",{"Source.Name"})

Order ID	Customer ID	Cookies Shipped	Revenue	Cost
1	413383	152689 - YT Restaurants	936	2340
2	751142	985245 - Quick Bite Convenience Stores	987	2467.5
3	876056	985245 - Quick Bite Convenience Stores	270	675
4	548247	985245 - Quick Bite Convenience Stores	925	2312.5
5	626350	152689 - YT Restaurants	974	2435
6	XXXXXX	XXXXXX - Promo	715	1787.5
7	200181	325698 - Cascade Grovers	454	1135
8	711368	152689 - YT Restaurants	484	1210
9	958419	325698 - Cascade Grovers	269	672.5
10	808349	985245 - Quick Bite Convenience Stores	396	990
11	392582	985245 - Quick Bite Convenience Stores	121	302.5
12	415147	152689 - YT Restaurants	373	932.5
13	656789	325698 - Cascade Grovers	118	295
14	247229	325698 - Cascade Grovers	137	342.5
15	532184	985245 - Quick Bite Convenience Stores	574	1435
16	412714	325698 - Cascade Grovers	541	1352.5
17	771432	325698 - Cascade Grovers	932	2330
18	363695	785432 - Park & Eat LLC	172	430
19	395821	452584 - Acme Grocery Stores	647	1617.5
20	234191	325698 - Cascade Grovers	660	1650
21	584039	325698 - Cascade Grovers	907	2267.5
22	463601	325698 - Cascade Grovers	429	1072.5
23	826310	325698 - Cascade Grovers	645	1612.5

8 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 9:46 AM

Query Settings

PROPERTIES

Name power query

APPLIED STEPS

- Source
- Filtered Hidden Files1
- Invoke Custom Function1
- Renamed Columns1
- Removed Other Columns1
- Expanded Table Column1
- Changed Type
- Removed Columns

Next, the column Order ID looks good for the most of the part, but we see the row 6 has messy data i.e., XXXXXX. So we need to get rid of this. I will uncheck XXXXXX and click ok to do so.

power query - Power Query Editor

File Home Transform Add Column View

Close & Load Refresh Preview Advanced Editor Manage Query

Choose Columns Remove Columns Manage Columns

Keep Rows Remove Rows Reduce Rows

Sort Split Column Group By

Data Type: Any Use First Row as Headers Replace Values

Merge Queries Append Queries Combine Files

Manage Parameters Combine Parameters

Data source settings Data Sources

New Source Recent Sources Enter Data New Query

Queries [5]

- Transform File from...
- Helper Queries [3]
 - Parameter1 (Sa...
 - Sample File
 - Transform File
 - Transform Sampl...
- Other Queries [1]
 - power query

fx = Table.SelectRows(#"Removed Columns", each ([Order ID] <> "XXXXXX"))

Order ID	Customer ID	Cookies Shipped	Revenue	Cost
1	413383	152689 - YT Restaurants	936	2340
2	751142	985245 - Quick Bite Convenience Stores	987	2467.5
3	876056	985245 - Quick Bite Convenience Stores	270	675
4	548247	985245 - Quick Bite Convenience Stores	925	2312.5
5	626350	152689 - YT Restaurants	974	2435
6	200181	325698 - Cascade Grovers	454	1135
7	711368	152689 - YT Restaurants	484	1210
8	958419	325698 - Cascade Grovers	269	672.5
9	808349	985245 - Quick Bite Convenience Stores	396	990
10	392582	985245 - Quick Bite Convenience Stores	121	302.5
11	415147	152689 - YT Restaurants	373	932.5
12	656789	325698 - Cascade Grovers	118	295
13	247229	325698 - Cascade Grovers	137	342.5
14	532184	985245 - Quick Bite Convenience Stores	574	1435
15	412714	325698 - Cascade Grovers	541	1352.5
16	771432	325698 - Cascade Grovers	932	2330
17	363695	785432 - Park & Eat LLC	172	430
18	395821	452584 - Acme Grocery Stores	647	1617.5
19	234191	325698 - Cascade Grovers	660	1650
20	584039	325698 - Cascade Grovers	907	2267.5
21	463601	325698 - Cascade Grovers	429	1072.5
22	826310	325698 - Cascade Grovers	645	1612.5
23	942932	985245 - Quick Bite Convenience Stores	260	1300

8 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 9:46 AM

Query Settings

PROPERTIES

Name power query

APPLIED STEPS

- Source
- Filtered Hidden Files1
- Invoke Custom Function1
- Renamed Columns1
- Removed Other Columns1
- Expanded Table Column1
- Changed Type
- Removed Columns
- Filtered Rows

Next, in Customer ID column, we see ID is combined with customer name, that doesn't look okay. So a better idea is to split these two information in separate columns.

power query - Power Query Editor

File Home Transform Add Column View

Close & Load Refresh Preview Advanced Editor Choose Columns Remove Columns Keep Rows Remove Rows Sort Split Column Group By Data Type: Any Use First Row as Headers Replace Values Merge Queries Append Queries Combine Files Manage Parameters Data source settings New Source Recent Sources Enter Data

Queries [5]

Transform File from... Helper Queries [3] Parameter1 (Sa... Sample File Transform File Transform Sampl... Other Queries [1] power query

Table.SplitColumn("#Filtered Rows", "Customer ID", Splitter.SplitTextByDelimiter("-", QuoteStyle.Csv),

Order ID	Customer ID.1	Customer ID.2	Cookies Shipped	Revenue	Cost
1	413383 152689	YT Restaurants	936	4680	
2	751142 985245	Quick Bite Convenience Stores	987	4935	
3	876056 985245	Quick Bite Convenience Stores	270	1350	
4	548247 985245	Quick Bite Convenience Stores	925	4625	
5	626350 152689	YT Restaurants	974	4870	
6	200181 325698	Cascade Grovers	454	2270	
7	711368 152689	YT Restaurants	484	2420	
8	958419 325698	Cascade Grovers	269	1345	
9	808349 985245	Quick Bite Convenience Stores	396	1980	
10	392582 985245	Quick Bite Convenience Stores	121	605	
11	415147 152689	YT Restaurants	373	1865	
12	656789 325698	Cascade Grovers	118	590	
13	247229 325698	Cascade Grovers	137	685	
14	532184 985245	Quick Bite Convenience Stores	574	2870	
15	412714 325698	Cascade Grovers	541	2705	
16	771432 325698	Cascade Grovers	932	4660	
17	363695 785432	Park & Eat LLC	172	860	
18	395821 452584	Acme Grocery Stores	647	3235	
19	234191 325698	Cascade Grovers	660	3300	
20	584039 325698	Cascade Grovers	907	4535	
21	463601 325698	Cascade Grovers	429	2145	
22	826310 325698	Cascade Grovers	645	3225	
23	942932 985245	Quick Bite Convenience Stores	260	1300	

Query Settings

PROPERTIES

Name

power query

APPLIED STEPS

Source

Filtered Hidden Files1

Invoke Custom Function1

Renamed Columns1

Removed Other Columns1

Expanded Table Column1

Changed Type

Removed Columns

Filtered Rows

Split Column by Delimiter

9 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 11:03 AM

We will rename these columns now.

power query - Power Query Editor

File Home Transform Add Column View

Close & Load Refresh Preview Advanced Editor Choose Columns Remove Columns Keep Rows Remove Rows Sort Split Column Group By Data Type: Text Use First Row as Headers Replace Values Merge Queries Append Queries Combine Files Manage Parameters Data source settings New Source Recent Sources Enter Data

Queries [5]

Transform File from... Helper Queries [3] Parameter1 (Sa... Sample File Transform File Transform Sampl... Other Queries [1] power query

Table.RenameColumns("#Split Column by Delimiter",{"Customer ID.1", "Customer ID"}, {"Customer ID.2", "Customer

Order ID	Customer ID	Customer Name	Cookies Shipped	Revenue	Cost
1	413383	152689 YT Restaurants	936	4680	
2	751142	985245 Quick Bite Convenience Stores	987	4935	
3	876056	985245 Quick Bite Convenience Stores	270	1350	
4	548247	985245 Quick Bite Convenience Stores	925	4625	
5	626350	152689 YT Restaurants	974	4870	
6	200181	325698 Cascade Grovers	454	2270	
7	711368	152689 YT Restaurants	484	2420	
8	958419	325698 Cascade Grovers	269	1345	
9	808349	985245 Quick Bite Convenience Stores	396	1980	
10	392582	985245 Quick Bite Convenience Stores	121	605	
11	415147	152689 YT Restaurants	373	1865	
12	656789	325698 Cascade Grovers	118	590	
13	247229	325698 Cascade Grovers	137	685	
14	532184	985245 Quick Bite Convenience Stores	574	2870	
15	412714	325698 Cascade Grovers	541	2705	
16	771432	325698 Cascade Grovers	932	4660	
17	363695	785432 Park & Eat LLC	172	860	
18	395821	452584 Acme Grocery Stores	647	3235	
19	234191	325698 Cascade Grovers	660	3300	
20	584039	325698 Cascade Grovers	907	4535	
21	463601	325698 Cascade Grovers	429	2145	
22	826310	325698 Cascade Grovers	645	3225	
23	942932	985245 Quick Bite Convenience Stores	260	1300	

Query Settings

PROPERTIES

Name

power query

APPLIED STEPS

Source

Filtered Hidden Files1

Invoke Custom Function1

Renamed Columns1

Removed Other Columns1

Expanded Table Column1

Changed Type

Removed Columns

Filtered Rows

Split Column by Delimiter

Renamed Columns

9 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 11:03 AM

Now, I wish to calculate the profit.

Profit = Revenue – Cost

For this, I selected both columns: revenue and cost, then headed over to add columns in the ribbon and chose standard that helped me subtract the revenue column with cost column and create a new column that I named Profit.

power query - Power Query Editor

File Home Transform Add Column View

Column From Examples Custom Invoke Custom Function Conditional Column Index Column Duplicate Column

Format Merge Columns ABC 123 Extract Parse

Statistics Standard Scientific 10² Trigonometry Rounding Information

Date Time Duration

General From Text From Number From Date & Time

Table.ReorderColumns(#"Renamed Columns2",{ "Order ID", "Customer ID", "Customer Name", "Cookies Shipped", "Revenue", "Cost", "Profit", "Order Date", "Ship Date" })

	1 ² Cookies Shipped	1 ² Revenue	1.2 Cost	1.2 Profit	Order Date	Ship Date	Order
1	936	4680	2340	2340	12/31/2017	1/5/2018	Shippec
2	987	4935	2467.5	2467.5	12/31/2017	1/4/2018	Shippec
3	270	1350	675	675	12/30/2017	1/3/2018	Shippec
4	925	4625	2312.5	2312.5	12/30/2017	1/3/2018	Shippec
5	974	4870	2435	2435	12/29/2017	1/3/2018	Shippec
6	454	2270	1135	1135	12/30/2017	1/2/2018	Shippec
7	484	2420	1210	1210	12/28/2017	1/2/2018	Shippec
8	269	1345	672.5	672.5	12/29/2017	1/1/2018	Shippec
9	396	1980	990	990	12/29/2017	1/1/2018	Shippec
10	121	605	302.5	302.5	12/28/2017	1/1/2018	Shippec
11	373	1865	932.5	932.5	12/27/2017	1/1/2018	Shippec
12	118	590	295	295	12/28/2017	12/30/2017	Shippec
13	137	685	342.5	342.5	12/27/2017	12/30/2017	Shippec
14	574	2870	1435	1435	12/26/2017	12/30/2017	Shippec
15	541	2705	1352.5	1352.5	12/27/2017	12/29/2017	Shippec
16	932	4660	2330	2330	12/27/2017	12/29/2017	Shippec
17	172	860	430	430	12/24/2017	12/29/2017	Shippec
18	647	3235	1617.5	1617.5	12/26/2017	12/28/2017	Shippec
19	660	3300	1650	1650	12/26/2017	12/28/2017	Shippec
20	907	4535	2267.5	2267.5	12/26/2017	12/28/2017	Shippec
21	429	2145	1072.5	1072.5	12/26/2017	12/28/2017	Shippec
22	645	3225	1612.5	1612.5	12/26/2017	12/28/2017	Shippec
23	260	1300	650	650	12/24/2017	12/27/2017	Shippec
24							

10 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 11:03 AM

Query Settings

PROPERTIES

Name

power query

APPLIED STEPS

Source

Filtered Hidden Files1

Invoke Custom Function1

Renamed Columns1

Removed Other Columns1

Expanded Table Column1

Changed Type

Removed Columns

Filtered Rows

Split Column by Delimiter

Renamed Columns

Inserted Subtraction

Renamed Columns2

X Reordered Columns

Note that my revenue, cost and profit are all currency, so I need to change the data type.

power query - Power Query Editor

File Home Transform Add Column View

Column From Custom Examples Column Invoke Custom Function Duplicate Column

Conditional Column Index Column

Format Parse

From Text

Statistics Standard Scientific

From Number

Trigonometry Rounding Information

Date Time Duration

From Date & Time

Queries

fx = Table.TransformColumnTypes(#"Reordered Columns",{"Profit", Currency.Type}, {"Cost", Currency.Type}, {"Revenue", Currency.Type})

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
		Cookies Shipped	\$ Revenue	\$ Cost	\$ Profit	Order Date	Ship Date	Order																
1		936	4,680.00		2,340.00	12/31/2017	1/5/2018	Shipped																
2		987	4,935.00		2,467.50	12/31/2017	1/4/2018	Shipped																
3		270	1,350.00		675.00	12/30/2017	1/3/2018	Shipped																
4		925	4,625.00		2,312.50	12/30/2017	1/3/2018	Shipped																
5		974	4,870.00		2,435.00	12/29/2017	1/3/2018	Shipped																
6		454	2,270.00		1,135.00	12/30/2017	1/2/2018	Shipped																
7		484	2,420.00		1,210.00	12/28/2017	1/2/2018	Shipped																
8		269	1,345.00		672.50	12/29/2017	1/1/2018	Shipped																
9		396	1,980.00		990.00	12/29/2017	1/1/2018	Shipped																
10		121	605.00		302.50	12/28/2017	1/1/2018	Shipped																
11		373	1,865.00		932.50	12/27/2017	1/1/2018	Shipped																
12		118	590.00		295.00	12/28/2017	12/30/2017	Shipped																
13		137	685.00		342.50	12/27/2017	12/30/2017	Shipped																
14		574	2,870.00		1,435.00	12/26/2017	12/30/2017	Shipped																
15		541	2,705.00		1,352.50	12/27/2017	12/29/2017	Shipped																
16		932	4,660.00		2,330.00	12/27/2017	12/29/2017	Shipped																
17		172	860.00		430.00	12/24/2017	12/29/2017	Shipped																
18		647	3,235.00		1,617.50	12/26/2017	12/28/2017	Shipped																
19		660	3,300.00		1,650.00	12/26/2017	12/28/2017	Shipped																
20		907	4,535.00		2,267.50	12/26/2017	12/28/2017	Shipped																
21		429	2,145.00		1,072.50	12/26/2017	12/28/2017	Shipped																
22		645	3,225.00		1,612.50	12/26/2017	12/28/2017	Shipped																
23		260	1,300.00		650.00	12/24/2017	12/27/2017	Shipped																

Query Settings

PROPERTIES

Name

power query

APPLIED STEPS

Source

Filtered Hidden Files1

Invoke Custom Function1

Renamed Columns1

Removed Other Columns1

Expanded Table Column1

Changed Type

Removed Columns

Filtered Rows

Split Column by Delimiter

Renamed Columns

Inserted Subtraction

Renamed Columns2

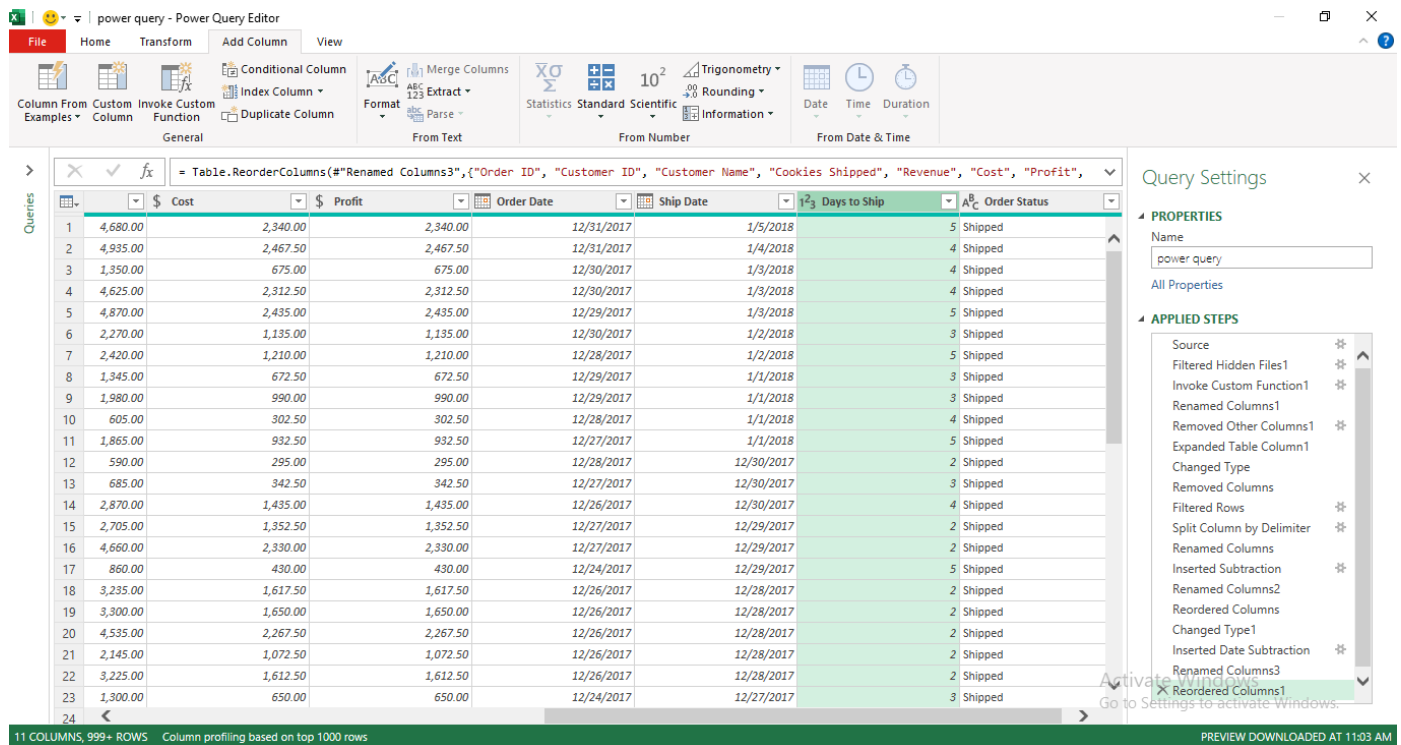
Reordered Columns

Changed Type1

10 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 11:03 AM

Next, we know that there is some, unfortunately, lag between the actual order date and shipped date. So how much was the lag, we need to calculate this. We can subtract dates by selecting shipped date first then order date and going to add column date option to subtract. We named our new column Days to Ship that tells us how many days it took for the company to ship the cookies from warehouse to the customer address.



After transforming the data, I closed the query and loaded it. I opened the file in Power BI Desktop to analyse and visualize my data. Here are some interesting results:

