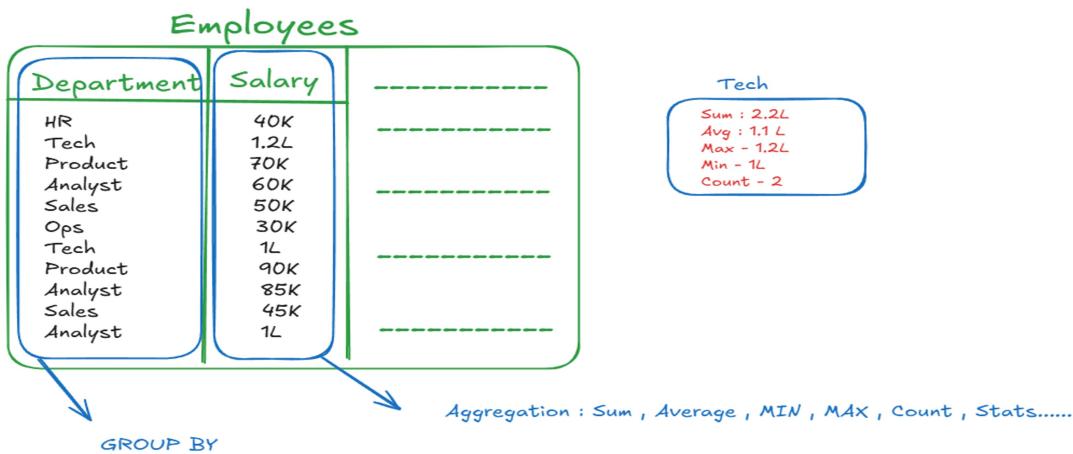


Power BI - Data Transformation - p3 - Lecture 4

Power BI - Data Transformation - p3



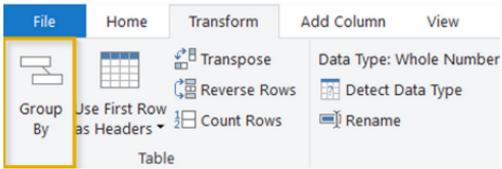
Queries [10]

ProductKey	Total Order Quantity
1	332
2	312
3	350
4	338
5	310
6	314
7	345
8	313
9	351
10	344
11	326
12	348
13	311
14	324
15	342
16	349

ProductKey	Total Order Quantity
1	385
2	383
3	326
4	352
5	369
6	342
7	375
8	356
9	360
10	324
11	330
12	368
13	362
14	340
15	334
16	373
17	358
18	377
19	332
20	379
21	371

ProductKey	Total Order Quantity
1	529
2	214
3	540
4	377
5	215
6	229
7	528
8	536
9	530
10	223
11	538
12	584
13	485
14	477
15	479
16	488
17	580
18	583
19	220
20	480
21	484

GROUPING & AGGREGATING



Group By allows you to aggregate data at a different level (i.e. group daily records into monthly, aggregate transactions by store, etc.)

	Order Date	Product Key	Customer Key	Order Quantity
1	6/25/2022	214	14719	1
2	10/8/2021	214	21990	1
3	12/30/2021	214	22098	1
4	6/19/2022	214	22748	1
5	8/16/2021	214	27821	1
6	10/9/2021	214	15685	1
7	8/9/2021	214	14951	1
8	1/19/2022	214	23101	1
9	9/23/2021	214	17158	2
10	1/19/2022	214	24196	1
11	6/29/2022	214	12963	1
12	9/13/2021	214	12715	1
13	10/2/2021	214	14846	1
14	7/31/2021	214	11290	1
15	11/24/2021	214	22103	1
16	8/1/2021	214	16982	1
17	10/12/2021	214	20410	1
18	9/10/2021	214	14217	1
19	10/22/2021	214	19642	1
20	8/11/2021	214	11666	1

Group By

Specify the column to group by and the desired output.

Basic Advanced

Product Key

New column name

TotalQuantity

Operation

Sum

Column

Order Quantity

OK

Cancel

	Product Key	TotalQuantity
1	214	2099
2	215	1940
3	220	1995
4	223	4151
5	226	392
6	229	408
7	232	424
8	235	381
9	310	169
10	311	139
11	312	179
12	313	168
13	314	157
14	320	65
15	322	39
16	324	72
17	326	65

Here we're transforming a daily transaction-level table into a summary of Total Quantity by Product Key

NOTE: Any fields not specified in the Group By settings are lost

The screenshot shows a table of daily transaction data with columns: Order Date, Product Key, Customer Key, and Order Quantity. An arrow points from the 'Group By' dialog box to the last row of the table, indicating the transformation process.

	Order Date	Product Key	Customer Key	Order Quantity
1	6/25/2022	214	14719	1
2	10/8/2021	214	21990	1
3	12/30/2021	214	22098	1
4	6/29/2022	214	22748	1
5	8/16/2021	214	27821	1
6	10/9/2021	214	15685	1
7	8/9/2021	214	14951	1
8	1/19/2022	214	23101	1
9	9/23/2021	214	17158	1
10	1/19/2022	214	24196	1
11	6/29/2022	214	12963	1
12	9/13/2021	214	12715	1
13	10/2/2021	214	14846	1
14	7/31/2021	214	11290	1
15	11/24/2021	214	22103	1
16	8/1/2021	214	16982	1
17	10/12/2021	214	20410	1
18	9/10/2021	214	14217	1
19	10/22/2021	214	19642	1
20	8/11/2021	214	11666	1

The screenshot shows the resulting summary table with columns: Product Key, Customer Key, and TotalQuantity. A green arrow points from the original table's Order Quantity column to the TotalQuantity column in the summary table, indicating the transformation.

	Product Key	Customer Key	1.2	TotalQuantity
1	214	19356	1	
2	214	15101	1	
3	214	12473	1	
4	214	12963	1	
5	214	26986	1	
6	214	13202	1	
7	214	14951	1	
8	214	11201	1	
9	214	19538	1	
10	214	22749	1	
11	214	15815	1	
12	214	19252	1	
13	214	14849	1	
14	214	11290	1	
15	214	27851	1	
16	214	16982	1	
17	214	21863	1	
18	214	19725	1	
19	214	15684	1	
20	214	11666	1	
21	214	26941	1	

This time we're transforming the daily, transaction-level table into a summary of Total Quantity grouped by both Product Key and Customer Key (using the "Advanced" option)

NOTE: This is like creating a PivotTable in Excel and pulling in Sum of Order Quantity with Product Key and Customer Key as row labels

\sum TerritoryKey	\sum ProductKey	\sum Total Orders
Valid Error Empty	100% 0% 0%	100% 0% 0%
7 distinct, 7 unique	1 distinct, 0 unique	6 distinct, 5 unique
4	312	47
10	312	12
9	312	51
7	312	10
1	312	24
6	312	25
8	312	10

PIVOTING & UNPIVOTING

Pivoting describes the process of turning distinct row values into columns, and unpivoting describes the process of turning distinct columns into rows

The diagram illustrates the relationship between two tables. The top table is a pivoted version of the bottom table. Arrows show the mapping from the bottom table's rows to the top table's columns.

Top Table (Pivoted):

Date	Product Category	North Region	Central Region	South Region
7/1/2022	Bikes	10	19	25
7/1/2022	Components	14	31	16
7/1/2022	Clothing	35	32	46

Bottom Table (Unpivoted):

Date	Product Category	Region	Quantity Sold
7/1/2022	Bikes	North Region	10
7/1/2022	Bikes	Central Region	19
7/1/2022	Bikes	South Region	25
7/1/2022	Components	North Region	14
7/1/2022	Components	Central Region	31
7/1/2022	Components	South Region	16
7/1/2022	Clothing	North Region	35
7/1/2022	Clothing	Central Region	32
7/1/2022	Clothing	South Region	46

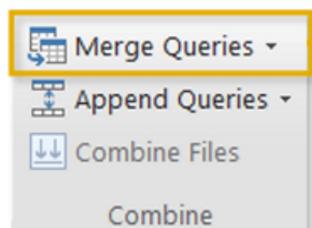
Annotations:

- An arrow labeled "PIVOTING" points from the bottom table to the top table.
- An arrow labeled "UNPIVOTING" points from the top table back to the bottom table.

Imagine the table on a hinge; pivoting rotates it from vertical to horizontal, and unpivoting rotates it from horizontal to vertical

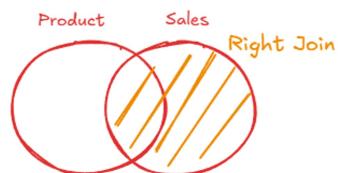
NOTE: Transpose works very similarly, but doesn't recognize unique values; instead, the entire table is transformed so that each row becomes a column and vice versa

MERGING QUERIES



Rule of Normalization.

-No Redundant Data.



Merge

Select a table and matching columns to create a merged table.

Sales Data
left
Product Price

Order Date	Product Key	Customer Key	Order Quantity	Index	Stock Date	Order Number	Territory
6/25/2022	214	14719	1	55115	4/20/2022	S073780	
10/8/2021	214	21990	1	14247	7/2/2021	S055746	
12/30/2021	214	22098	1	26322	11/10/2021	S061052	
6/29/2022	214	22748	1	55740	4/9/2022	S074069	

Product Lookup
Right

Product Key	Product Subcategory Key	Product S K U	Product Name	Model Name
214	31	HL-US09-R	Sport-100 Helmet, Red	Sport-100 Universal fit, v
215	31	HL-US09	Sport-100 Helmet, Black	Sport-100 Universal fit, v
218	23	SO-B909-M	Mountain Bike Socks, M	Mountain Bike Socks Combination c
219	23	SO-B909-L	Mountain Bike Socks, L	Mountain Bike Socks Combination c

Join Kind: Left Outer (all from first, matching from second)

Use fuzzy matching to perform the merge

> Fuzzy matching options

✓ The selection matches 56046 of 56046 rows from the first table.

OK Cancel

Merging queries allows you to join tables based on a common column (like a lookup in Excel)

In this case we're merging the Sales Data table with the Product Lookup table, which share a common Product Key column

NOTE: Merging adds columns to an existing table/query

Just because you can merge tables, doesn't mean you should!

In many cases, it's better to keep tables separate and define relationships between them in the data model (more on that soon!)

Table.ColumnName

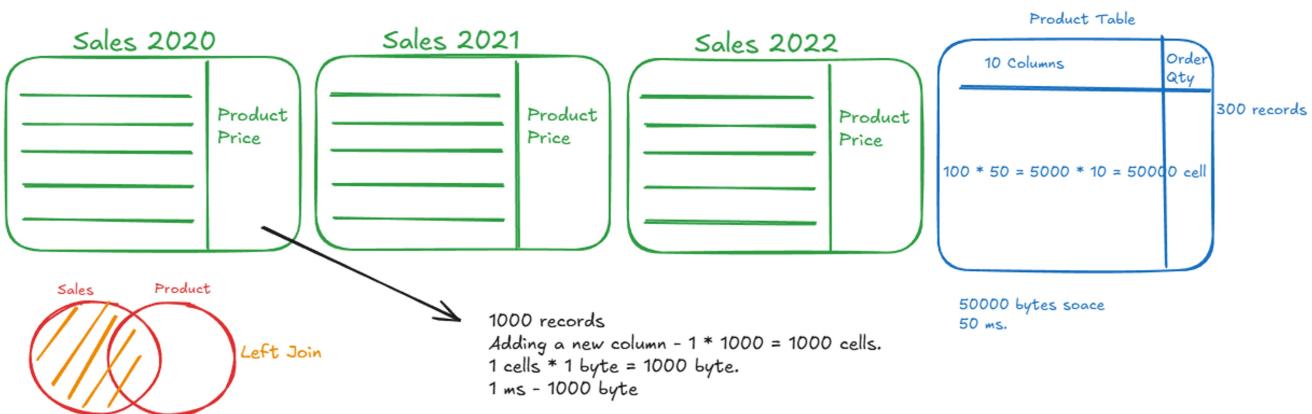
2.5 years of span

2020 - 2630

2021 - 23935

2022 - 29481

Product Table - 293

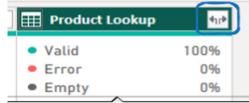


Merge

Select a table and matching columns to create a merged table.

Sales Data 2020

Index	OrderDate	StockDate	OrderNumber	ProductKey	CustomerKey	TerritoryKey	OrderLineItem
1	01-01-2020	21-09-2019	SO45080	332	14657	1	
2	01-01-2020	05-12-2019	SO45079	312	29255	4	
3	01-01-2020	29-10-2019	SO45082	350	11455	9	
4	01-01-2020	16-11-2019	SO45081	338	26782	6	



Merging - Horizontal Direction.

Join Kind: Left Outer (all from first, matching from second)

Use fuzzy matching to perform the merge

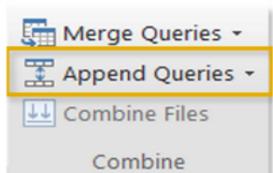
Fuzzy matching options

✓ The selection matches 2630 of 2630 rows from the first table.

OK Cancel

APPENDING QUERIES

Appending queries allows you to combine or stack tables sharing the exact same column structure and data types



Append

Concatenate rows from two tables into a single table.

Two tables Three or more tables

First table

AdventureWorks Sales Data 2020

Second table

AdventureWorks Sales Data 2021

V
E
R
T
I
C
A
L

Here we're appending the Adventure Works Sales 2020 table to the Adventure Works Sales 2021 table, which is valid since they share identical table structures.

NOTE: Appending adds rows to an existing table/query

- Sales Data 2020
- Sales Data 2021
- Sales Data 2022
- Product Category Sales (...)
- Sales Data 2020 - 2022

Append

Concatenate rows from three or more tables into a single table.

Two tables Three or more tables



REFRESHING QUERIES

The screenshot shows the Power BI ribbon with the 'Home' tab selected. In the 'External Tools' section, the 'Refresh' button is highlighted with a yellow box. Below the ribbon, the 'Query Editor' interface is shown. A list of queries is visible on the left, and a context menu is open over the 'Customer Lookup' query. The menu options include 'Copy', 'Paste', 'Delete', 'Rename', 'Enable load', 'Include in report refresh' (which is highlighted with a yellow box), 'Duplicate', 'Reference', 'Move To Group', 'Move Up', 'Move Down', 'Create Function...', 'Convert To Parameter', 'Advanced Editor', and 'Properties...'. A handwritten note on the left side of the screen says: '- By default, all queries will refresh when you use the Refresh command from the Home tab'.

From the Query Editor, uncheck Include in report refresh to exclude individual queries from the refresh

PRO TIP: Exclude queries from refresh that don't change often (like lookups or static data tables)

POWER QUERY BEST PRACTICES

Get organized before connecting and loading data

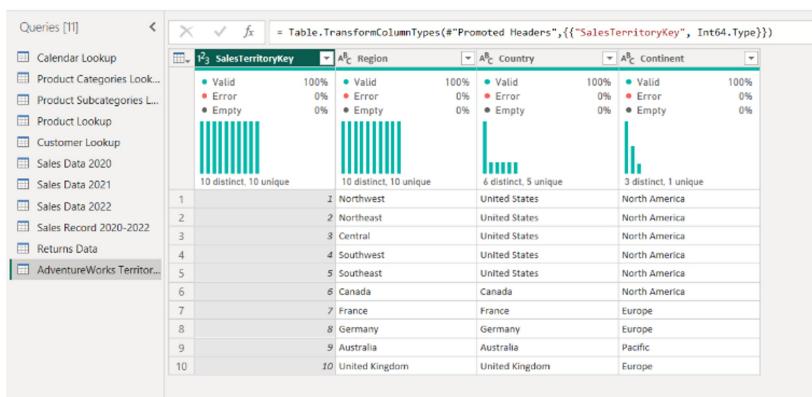
- Define clear and intuitive table/query names from the start, and establish an organized file/folder structure if you are working with local flat files to avoid changes to file names or paths.

Disable report refresh for any static data sources

- There's no need to constantly refresh data sources that don't change, like lookups or static data tables

When working with large tables, only load the data you need

- Don't include hourly data when you only need daily, or transaction-level data when only need a product-level summary (extra data will only slow your report down!)



Queries [11]

AdventureWorks-CX111

File Home Transform Add Column

Close & Apply New Source Recent Enter Data Data source settings Data Sources

Close New Query Data Part

Queries [11]

- Calendar
- Product
- Product
- Product
- Customer
- Sales Data Enable load Include in report refresh
- Sales Data Duplicate
- Sales Data Reference
- Sales Record
- Returns Data
- AdventureWorks

Advanced Editor Properties...

10
11
12

You will find the option to exclude Lookup table from refresh. Uncheck the Option : Include in report refresh.

NOTE : Only remove lookup table as they are having static information other than FACT TABLE[sales , Return].

Refresh

- Sales Data 2020
56.8 KB from AdventureWorks Product Lookup.csv
- Sales Data 2021
428 KB from AdventureWorks Sales Data 2021.csv
- Sales Data 2022
486 KB from AdventureWorks Sales Data 2022.csv
- Sales Record 2020-2022
178 KB from AdventureWorks Sales Data 2021.csv
- Returns Data
35.6 KB from AdventureWorks Returns Data.csv

Cancel