

Data Transformation - 2

29 August 2024 21:16

Data Transformation - 2

Sales Tables:

- Append sales data from 2015, 2016, and 2017 into a new file, arranging them in a vertical manner.

Appending

- Merging up different one below other and creating a new table.

id	Name	Price
Sales 2015		

id	Name	Price
Sales 2016		

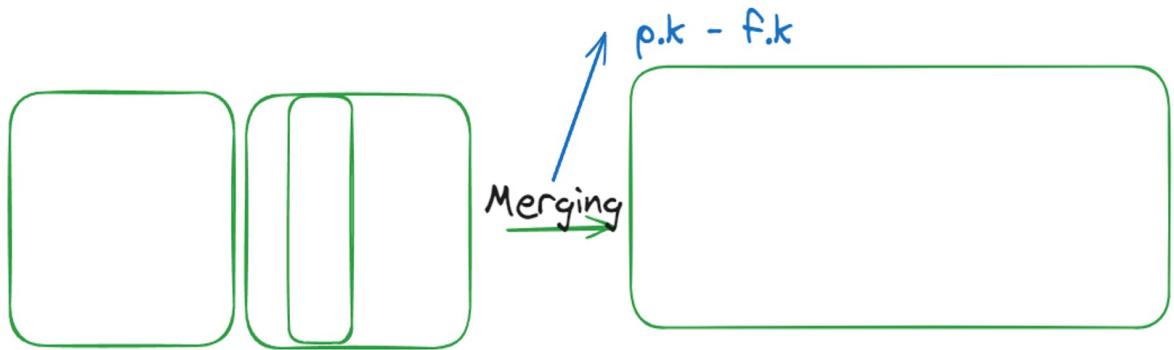
id	Name	Price
Sales 2017		

Sales 2015-17

Appending →

← Can't delete it as they are now reference table.

• o.k - f.k



Date $\xrightarrow{\text{error}}$ ABC \rightarrow Locale to convert into desired Date.

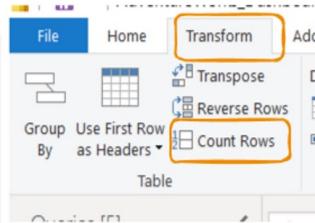
PROPERTIES

Name
sales 2015

All Properties

APPLIED STEPS

- Source
- Promoted Headers
- Changed Type with Locale
- Changed Type with Locale1
- Changed Type**



$$2630 + 23935 + 29481 = 56046$$

A

B

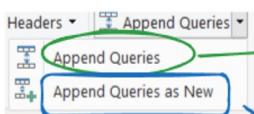
C

2.5 years

[Appended Table]

Appending : 2 options :

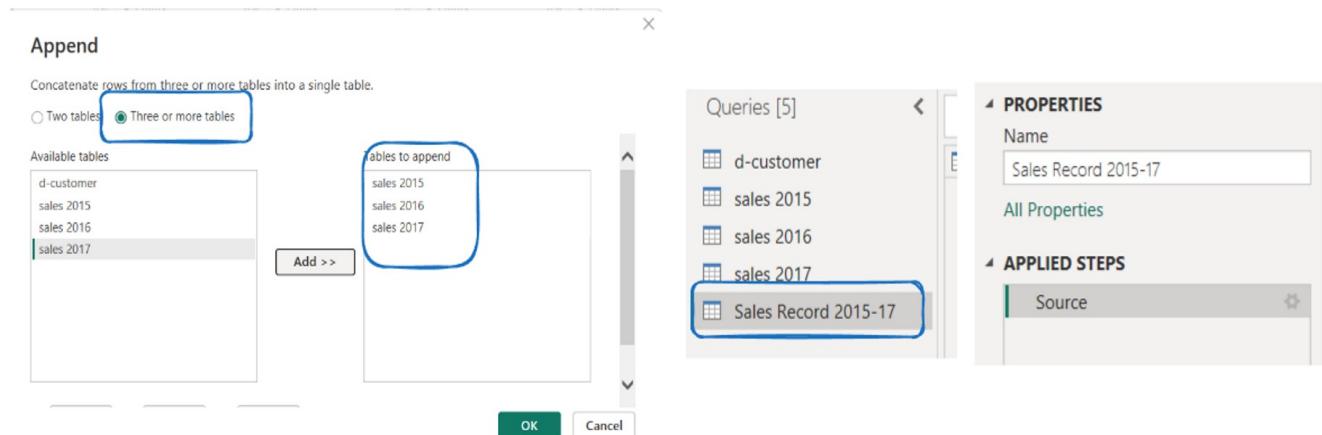
- Whether you want to append on one existing.
- Other option is append all the table and create the new table.



2015 , onwards start appending the data

It creates a new table to store individual

It creates a new table to store individual data.



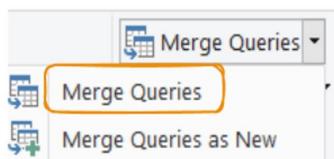
Product Tables:

- Merge the product price and product cost columns from the product table, gender, and occupation from the customer table into the sales table horizontally.
- In the merged table, group gender and occupation based on order quantity.



Step 1 : Import Product Data & Clean.

Step 2 : Merge Sales & Product to fetch 2 columns.



Merge

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

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

Sales Record 2015-17

OrderDate	StockDate	OrderNumber	ProductKey	CustomerKey	TerritoryKey	OrderLineItem	Order
01-01-2015	21-09-2001	SO45080	332	14657	1	1	
01-01-2015	05-12-2001	SO45079	312	29255	4	1	
01-01-2015	29-10-2001	SO45082	350	11455	9	1	
01-01-2015	16-11-2001	SO45081	338	26782	6	1	

d-product

ProductKey	ProductSubcategoryKey	ProductSKU	ProductName	ModelName	ProductDescription
214	31	HL-U509-R	Sport-100 Helmet, Red	Sport-100	Universal fit, well
215	31	HL-U509	Sport-100 Helmet, Black	Sport-100	Universal fit, well
218	23	SO-B909-M	Mountain Bike Socks, M	Mountain Bike Socks	Combination of n
219	23	SO-B909-L	Mountain Bike Socks, L	Mountain Bike Socks	Combination of n

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

OrderQuantity d-product

Search Columns to Expand A-Z

Expand Aggregate

(Select All Columns)

ProductKey

ProductSubcategoryKey

ProductSKU

ProductName

ModelName

ProductDescription

ProductColor

ProductSize

ProductStyle

ProductCost

ProductPrice

Use original column name as prefix

OK Cancel

\$ d-product.ProductCost \$ d-product.ProductPrice

Valid 100% Error 0% Empty 0%

Valid 100% Error 0% Empty 0%

4 distinct, 0 unique 4 distinct, 0 unique

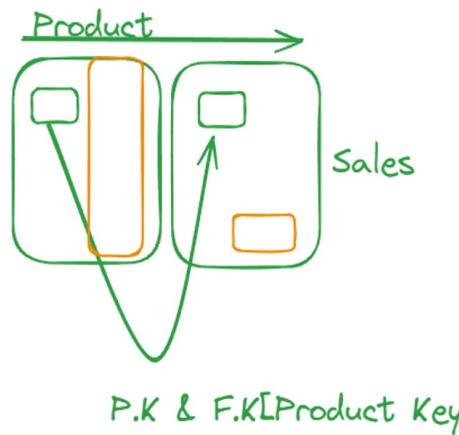
413.15 699.10

413.15 699.10

2,171.29 3,578.27

2,171.29 3,578.27

table.columnName



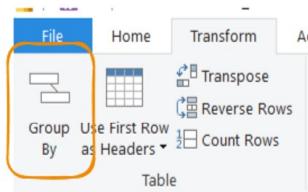
Same above steps for Customer Table Merged with Sales Table.

	A _B C	d-customer.Gender	A _B C	d-customer.Occupation
		Valid	100%	Valid
		Error	0%	Error
		Empty	0%	Empty
				
		2 distinct, 0 unique		5 distinct, 0 unique
7	F			Manual
7	F			Manual
7	F			Management
7	F			Professional
9	M			Management
9	M			Clerical
9	F			Management
9	F			Professional
9	M			Professional
7	M			Professional
7	M			Manual
7	M			Clerical
7	F			Professional
7	F			Professional
7	M			Clerical
7	M			Clerical
7	F			Skilled Manual
7	M			Clerical
9	M			Clerical
9	F			Professional

Product Tables:

- Merge the product price and product cost columns from the product table, gender, and occupation from the customer table into the sales table horizontally.
 - In the merged table, group gender and occupation based on order quantity.

- Merge the product price and product cost columns from the product table, gender, and occupation from the customer table into the sales table horizontally.
- In the merged table, group gender and occupation based on order quantity.



Group By

Specify the columns to group by and one or more outputs.

Basic Advanced

d-customer.Gender

d-customer.Occupation

Add grouping

New column name	Operation	Column
Total Order Quantity	Sum	OrderQuantity
Add aggregation		

OK Cancel

A ^B _C	d-customer.Gender	A ^B _C	d-customer.Occupation	1.2 Total Order Quantity
● Valid	91%	● Valid	91%	100%
● Error	0%	● Error	0%	0%
● Empty	9%	● Empty	9%	0%
3 distinct, 1 unique		6 distinct, 1 unique		11 distinct, 11 unique
F	Manual			2402
F	Management			4118
F	Professional			6769
M	Management			4448
M	Clerical			3083
M	Professional			6841
M	Manual			2462
F	Skilled Manual			5223
M	Skilled Manual			4793
F	Clerical			3076
	null		null	40959

Road-650 customer doesn't exist

312 29255

Road-150

M,F,Occupations.

I've to fix it.