

Model like a star - Star schema 101

Ynte Jan Kuindersma

THANK YOU

Data Community Austria Day 2025

Awesome Partner

Microsoft

Platinum

redgate

ACP IT for innovators.

Gold

b.telligent
smart data. smart decisions.

Lucient

Bronze

Tabular Editor

paiqo
The Platform & AI Company

Power BI Camp
www.linearis.at

Ynte Jan Kuindersma



- info@birdautomation.nl
- Groningen, Netherlands

- Freelance software developer
- MS SQL server, .NET, Excel, Power BI, VBA
- Microsoft Certified Data Analyst (DA-100)
- Microsoft Certified for Fabric (DP 600)
- Public speaker at Power Platform events

Why this Session?

- So many Power BI reports have too complex data structures
  - Data are taken as they come
  - People want to visualize as quickly as possible, not think about technical stuff
- And then the reports get slow and the DAX complicated

What do you see?

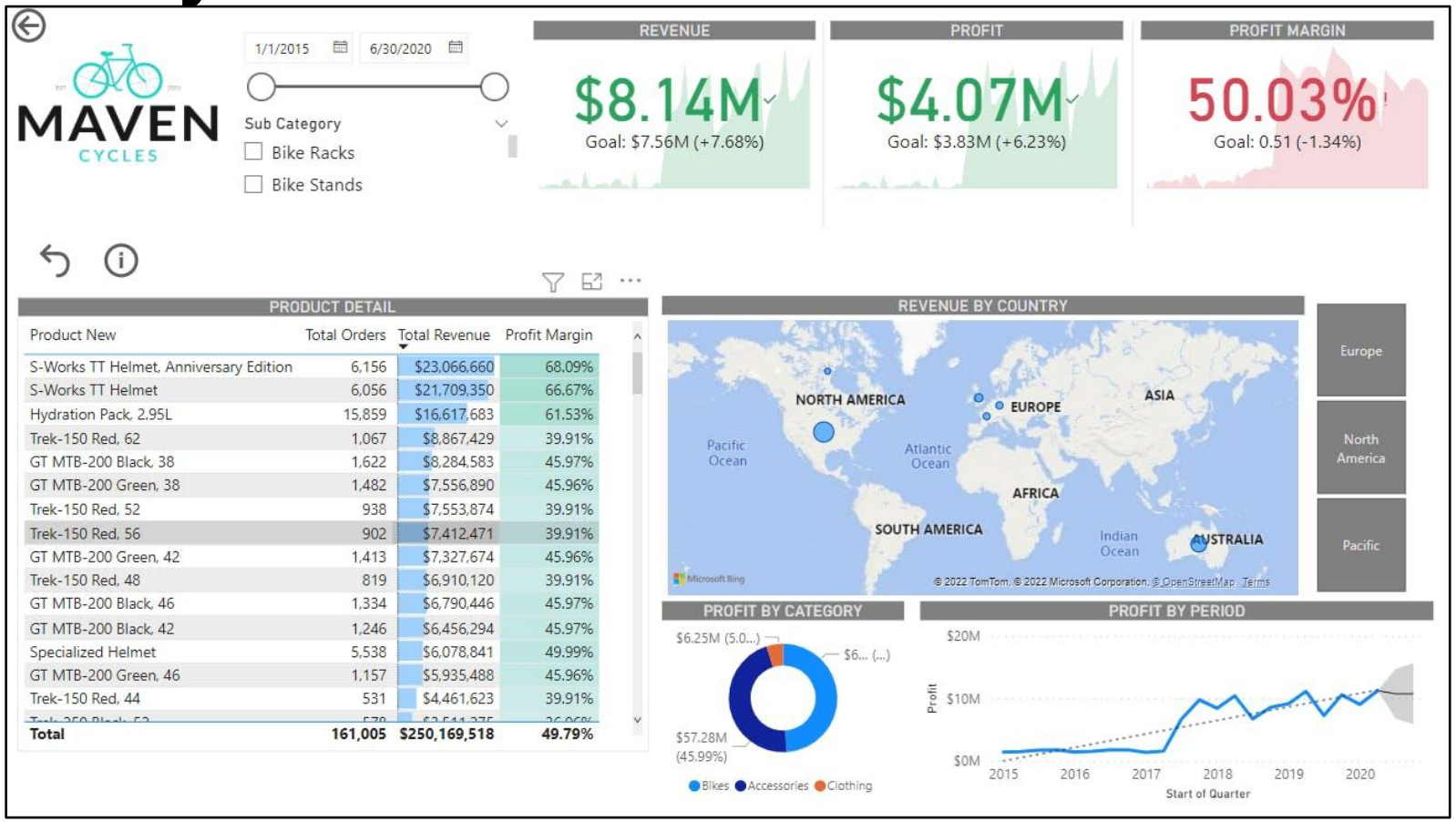


How do you get there?

- You want a well-prepared meal,
- so put some effort in getting the right ingredients
- And wash, chop, cook, spice etc them



What do you see?



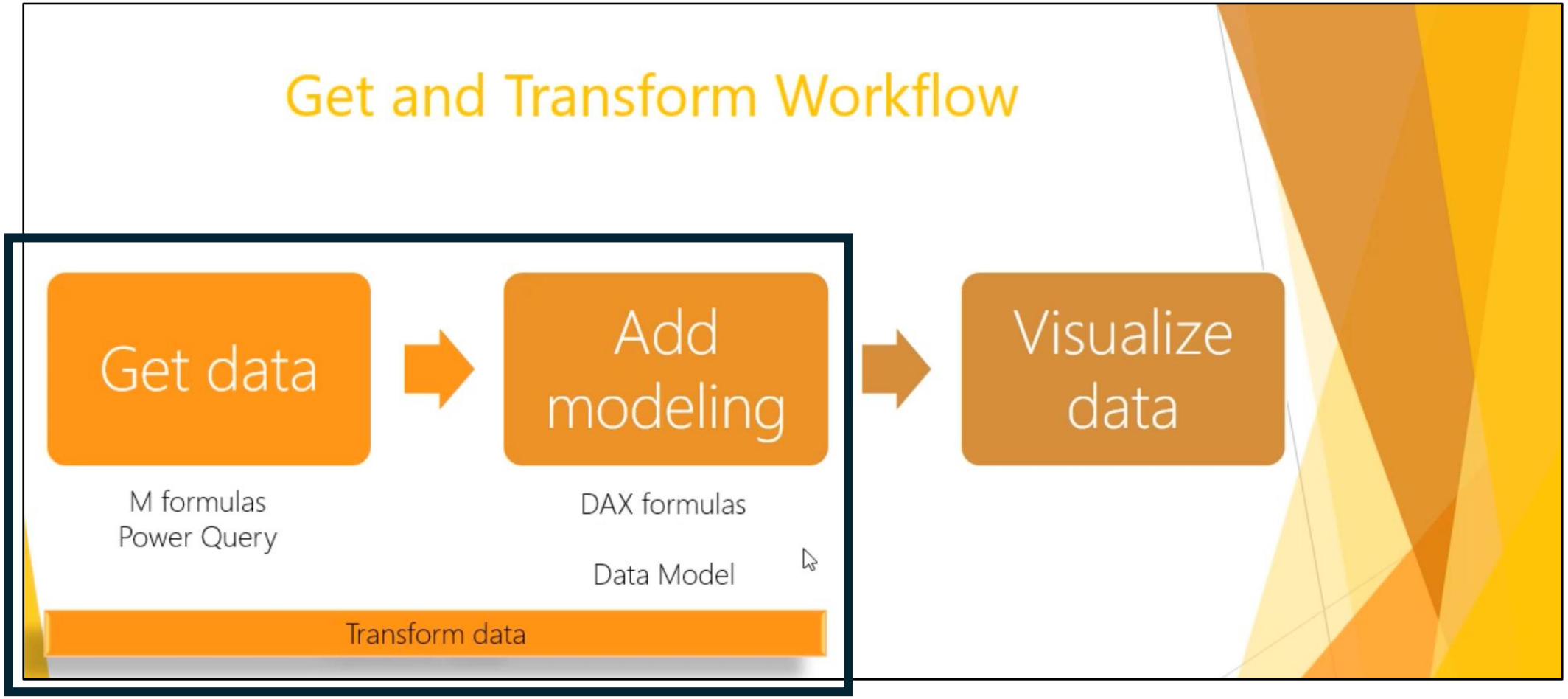
https://www.novypro.com/project/power-bi-project-maven-cycles

How do you get there?

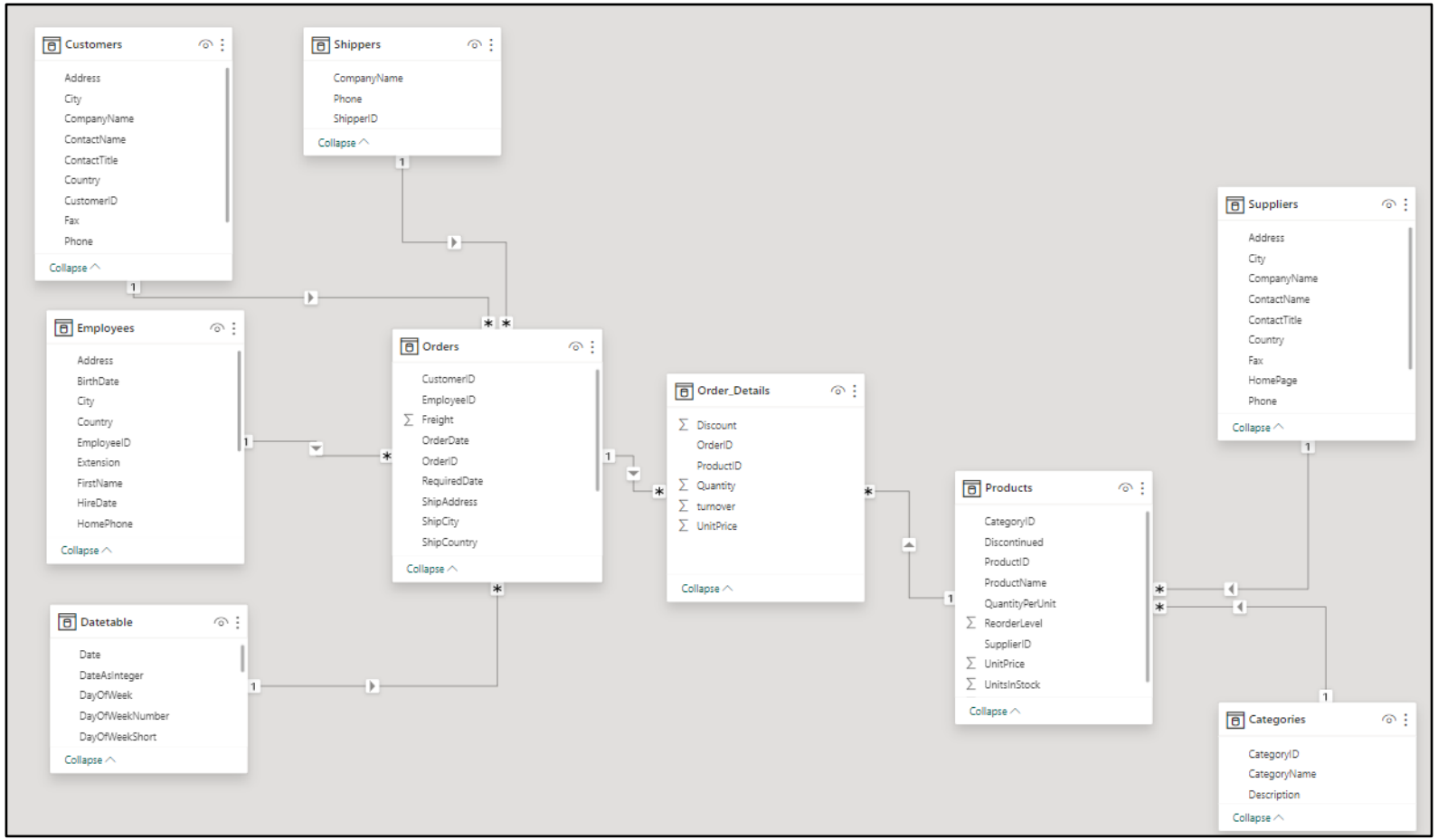
- You want a well-prepared report,
- so put some effort in getting the right data
- And wash, chop, cook, spice etc. them till they fit your purpose
- Invest!



Power BI Desktop report workflow



Real world example before Modelling – Normalized SQL server tables



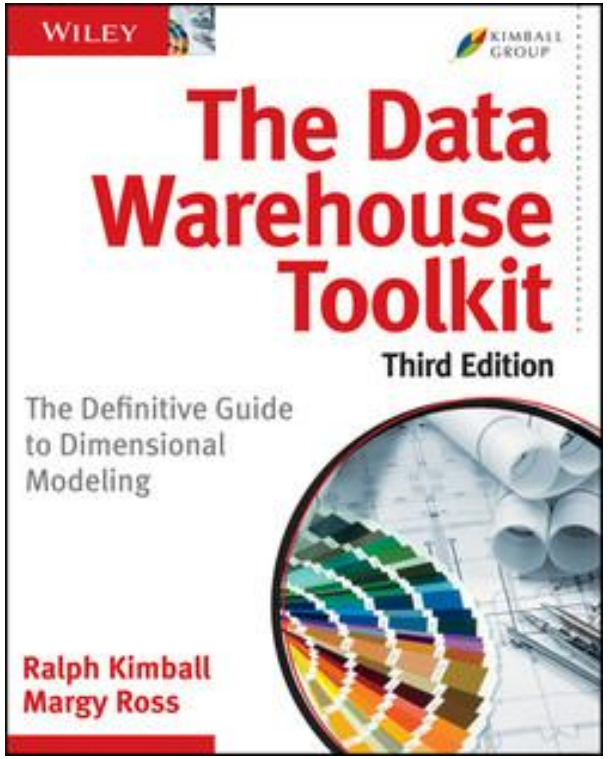
Real world example before Modelling - OneBigTable

	A	B	C	D	E	F	G
1	**CompanyName**	**LastName**	**OrderID**	**OrderDate**	**ProductName**	**Quantity**	**UnitPrice**
2	Blondesddsl père et fils	Fuller	10265	25-7-1996 00:00	Outback Lager	20	12
3	Blondesddsl père et fils	Fuller	10265	25-7-1996 00:00	Alice Mutton	30	31.2
4	Centro comercial Moctezuma	Peacock	10259	18-7-1996 00:00	Gravad lax	1	20.8
5	Centro comercial Moctezuma	Peacock	10259	18-7-1996 00:00	Sir Rodney's Scones	10	8
6	Chop-suey Chinese	Buchanan	10254	11-7-1996 00:00	Guaraná Fantástica	15	3.6
7	Chop-suey Chinese	Buchanan	10254	11-7-1996 00:00	Longlife Tofu	21	8
8	Chop-suey Chinese	Buchanan	10254	11-7-1996 00:00	Pâté chinois	21	19.2
9	Ernst Handel	Davolio	10258	17-7-1996 00:00	Mascarpone Fabioli	6	25.6
10	Ernst Handel	Davolio	10258	17-7-1996 00:00	Chang	50	15.2
11	Ernst Handel	Davolio	10258	17-7-1996 00:00	Chef Anton's Gumbo Mix	65	17
12	Ernst Handel	Dodsworth	10263	23-7-1996 00:00	Guaraná Fantástica	28	3.6
13	Ernst Handel	Dodsworth	10263	23-7-1996 00:00	Longlife Tofu	36	8
14	Ernst Handel	Dodsworth	10263	23-7-1996 00:00	Pavlova	60	13.9
15	Ernst Handel	Dodsworth	10263	23-7-1996 00:00	Nord-Ost Matjeshering	60	20.7
16	Folk och fä HB	Suyama	10264	24-7-1996 00:00	Jack's New England Clam Chowder	25	7.7
17	Folk och fä HB	Suyama	10264	24-7-1996 00:00	Chang	35	15.2
18	Frankenversand	Peacock	10267	29-7-1996 00:00	Lakkalikööri	15	14.4
19	Frankenversand	Peacock	10267	29-7-1996 00:00	Boston Crab Meat	50	14.7

Star schema, what is it?

- Kimball, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling (2013).*

- Spoiler
  - create one Fact table per reporting area
  - Create reusable Dimension tables



Star schema, what is it?

- Fact tables contain
  - Fact fields = columns with numbers & dates. Normally used as data field for lines etc.
  - Keys fields = columns with IDs to connect to Dimensions

- Example: sales quantity, sales price, turnover, order date, cost price, productid, customerID

Star schema, what is it?

- Dimension tables contain
  - Fields that give context to your data
  - Normally used on the axis & in slicers

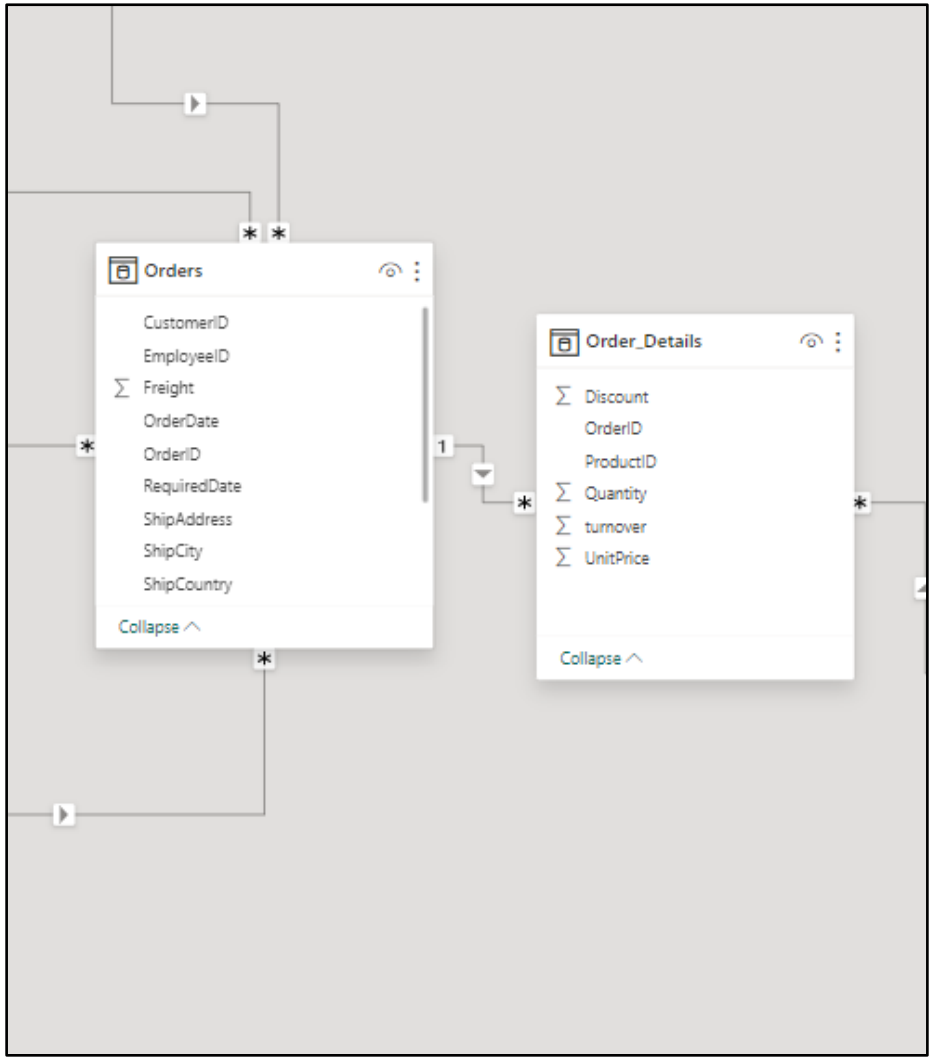
- Example: Products, Customers, Regions, Time

2 Problems you may run into...

- Different levels of data in a normalized datasource
  - Header / Details
  - Mulitple levels of relations
  - Solution: So this is too normalized: you need to **denormalize**

- OneBigTable (OBT)
  - Solution: This is not normalized at all: so you need to **normalize**

First demo


Orders
CustomerID
EmployeeID
Freight
OrderDate
OrderID
RequiredDate
ShipAddress
ShipCity
ShipCountry
Collapse
Order_Details
Discount
OrderID
ProductID
Quantity
turnover
UnitPrice
Collapse
1

Different levels of data?

- Merge!
- Merge tables to combine Fact columns and ID columns from X tables into one Fact table
  - This week I merged 10 tables into 1
- Eventually get extra Dimension IDs from other table(s)

Summary of our target: One Fact table

Facts	Dimensionkeys

• From Orders
  • Freight
  • OrderDate
• From Order_Details
  • Discount
  • Quantity
  • Turnover
  • UnitPrice
	• From Orders
	  • CustomerID
	  • EmplyeeID
	  • OrderID
	• From Order_Details
	  • ProductID
	• Extra: From Products
	  • CategoryID
	  • SupplierID

Target: One Fact table



CustomerID	EmployeeID	ProductID	OrderDate	Freight	UnitPrice	Quantity	Discount	turnover
ALFKI	6	28	25-8-1997 00:00:00	29,46	45,6	15	0,25	684
ALFKI	6	39	25-8-1997 00:00:00	29,46	18	21	0,25	378
ALFKI	6	46	25-8-1997 00:00:00	29,46	12	2	0,25	24
ALFKI	4	63	3-10-1997 00:00:00	61,02	43,9	20	0	878
ALFKI	4	3	13-10-1997 00:00:00	23,94	10	6	0	60
ALFKI	4	76	13-10-1997 00:00:00	23,94	18	15	0	270
ALFKI	1	59	15-1-1998 00:00:00	69,53	55	15	0	825
ALFKI	1	77	15-1-1998 00:00:00	69,53	13	2	0,2000000029	26
ALFKI	1	28	16-3-1998 00:00:00	40,42	45,6	2	0	91,2
ALFKI	1	6	16-3-1998 00:00:00	40,42	25	16	0,0500000007	400
ALFKI	3	71	9-4-1998 00:00:00	1,21	21,5	20	0	430
ALFKI	3	58	9-4-1998 00:00:00	1,21	13,25	40	0,0500000007	530

Demo time

- DEMO_1_factsales.pbix



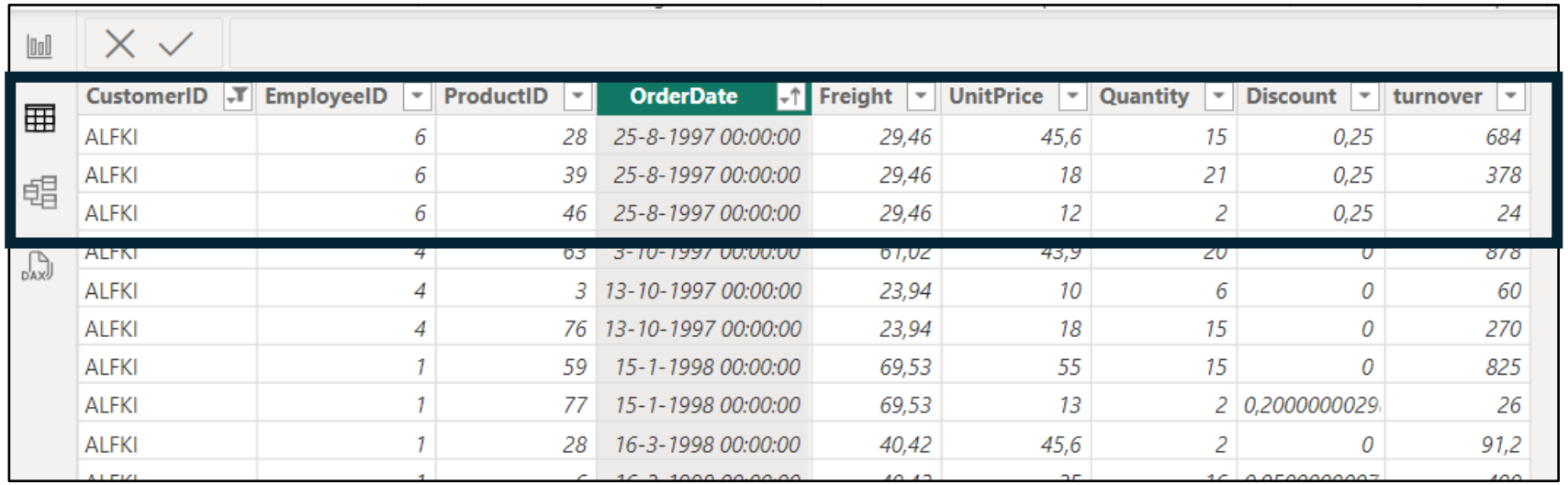
There is one new problem to solve

- Duplication of data fields?
  - Date
  - Freight

- Solution: You need to allocate Header data to detail level

Second demo

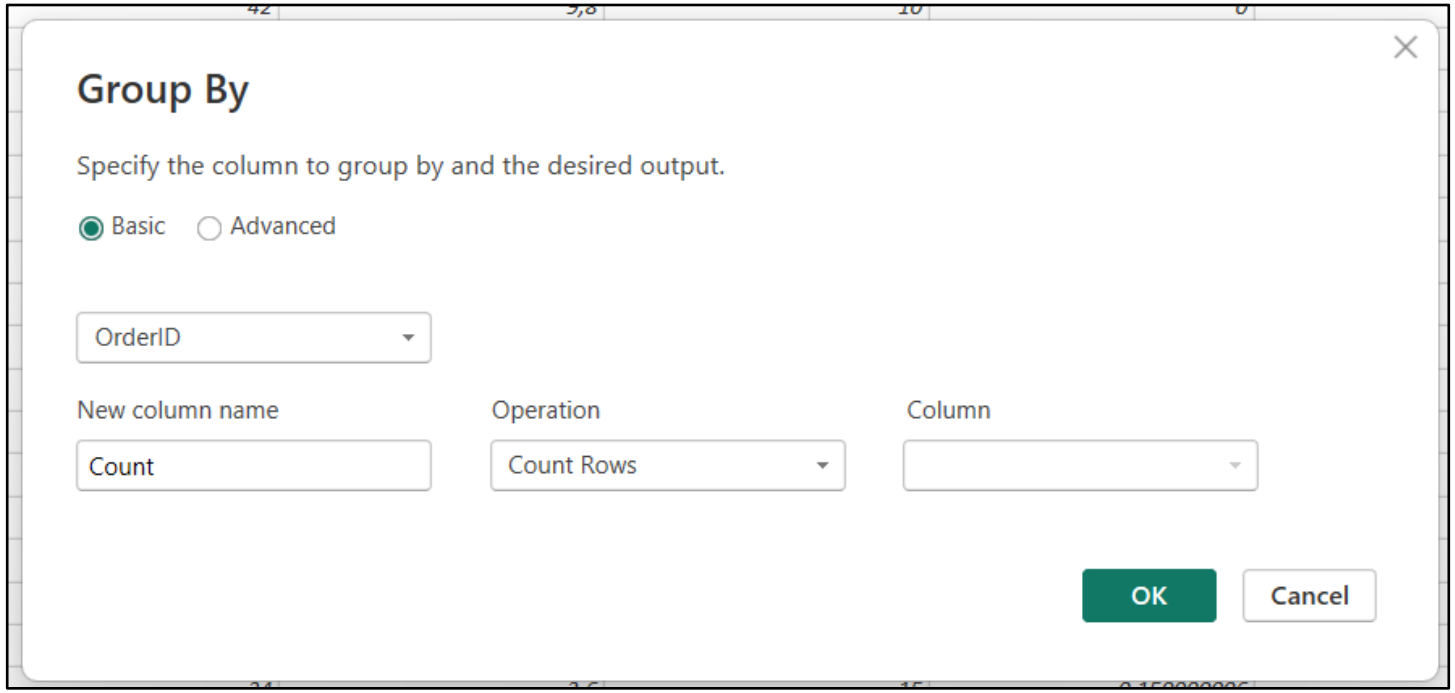
- Divide by the number of orderdetails lines?
  - 29,46 divided by 3
- Calculate row turnover as percentage of total Order turnover?
  - 29,46 multiplied with a factor: 684 divided by 1.086, etc



CustomerID	EmployeeID	ProductID	OrderDate	Freight	UnitPrice	Quantity	Discount	turnover
ALFKI	6	28	25-8-1997 00:00:00	29,46	45,6	15	0,25	684
ALFKI	6	39	25-8-1997 00:00:00	29,46	18	21	0,25	378
ALFKI	6	46	25-8-1997 00:00:00	29,46	12	2	0,25	24
ALFKI	4	63	3-10-1997 00:00:00	61,02	43,9	20	0	878
ALFKI	4	3	13-10-1997 00:00:00	23,94	10	6	0	60
ALFKI	4	76	13-10-1997 00:00:00	23,94	18	15	0	270
ALFKI	1	59	15-1-1998 00:00:00	69,53	55	15	0	825
ALFKI	1	77	15-1-1998 00:00:00	69,53	13	2	0,2000000029	26
ALFKI	1	28	16-3-1998 00:00:00	40,42	45,6	2	0	91,2

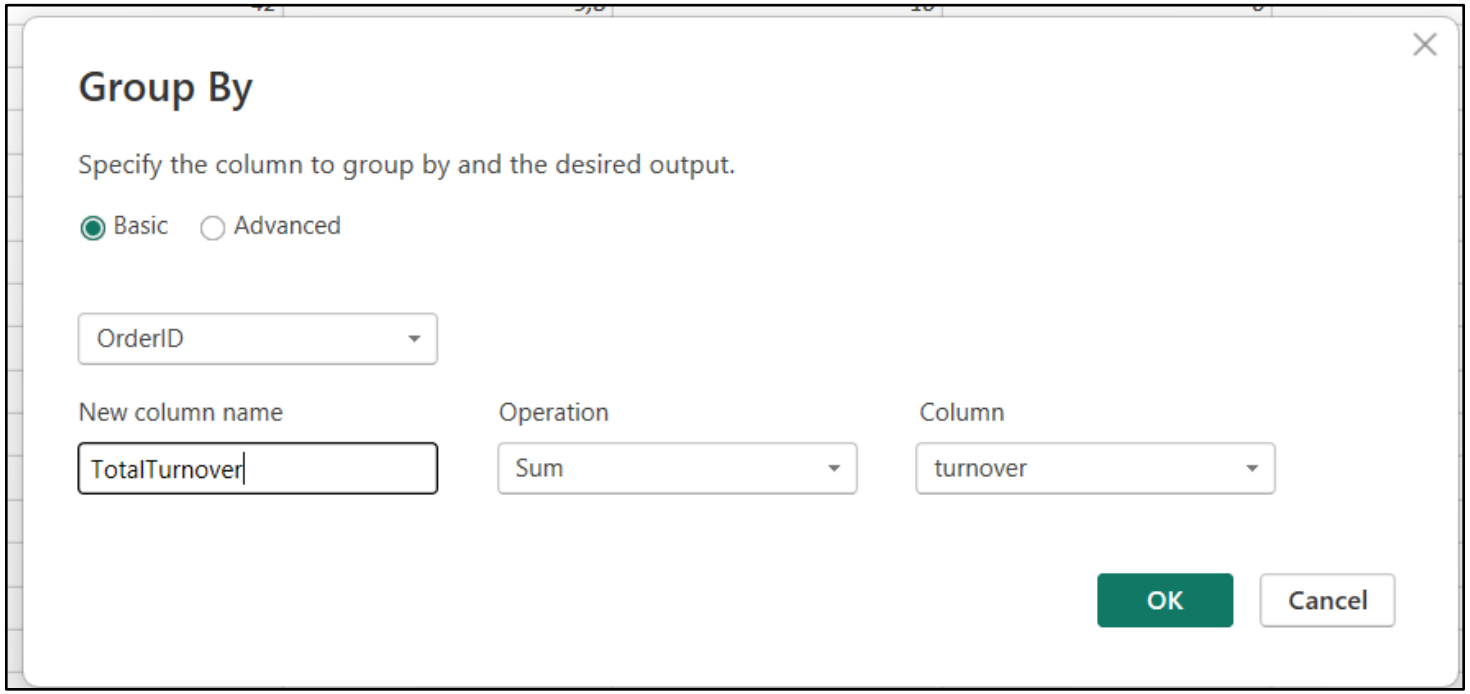
Allocate Header amount to the detail lines solution 1

- Divide by the number of detail lines
- Create helper table in Power Query with Group By & Count



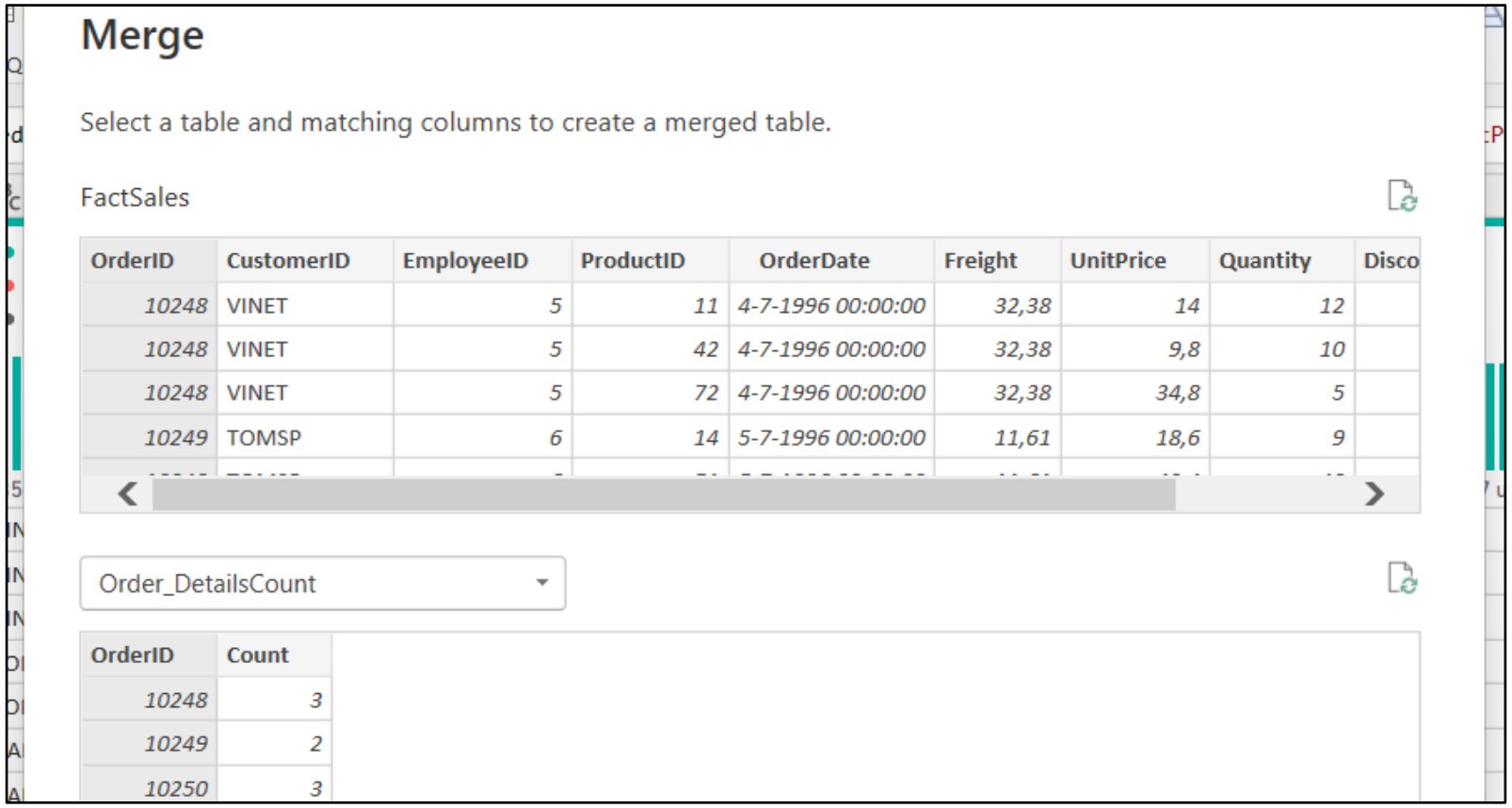
Allocate Header amount to the detail lines solution 2

- Calculate every rows turnover as percentage of order total?
- Create helper table in Power Query with Group By & Sum



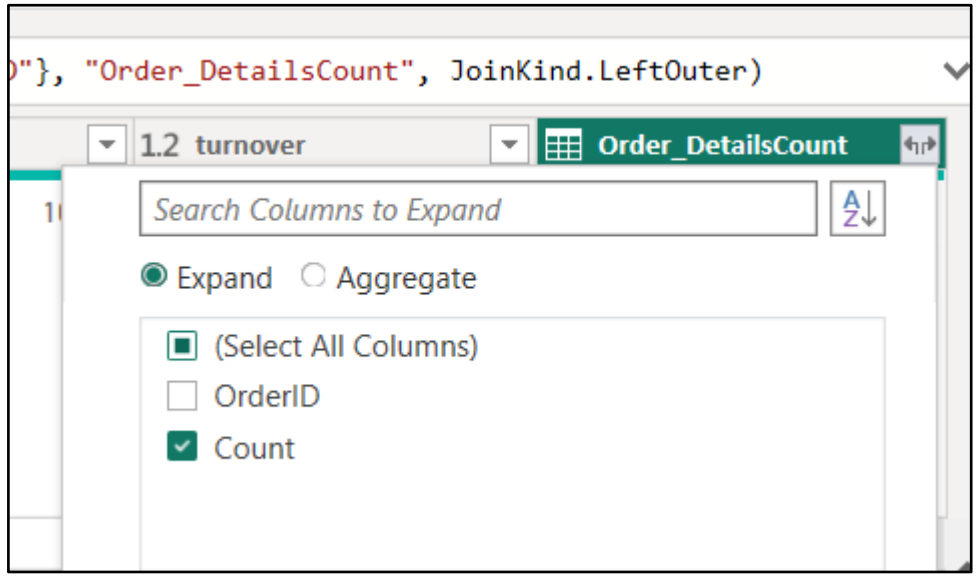
Allocate Header amount to the detail lines solution 1a

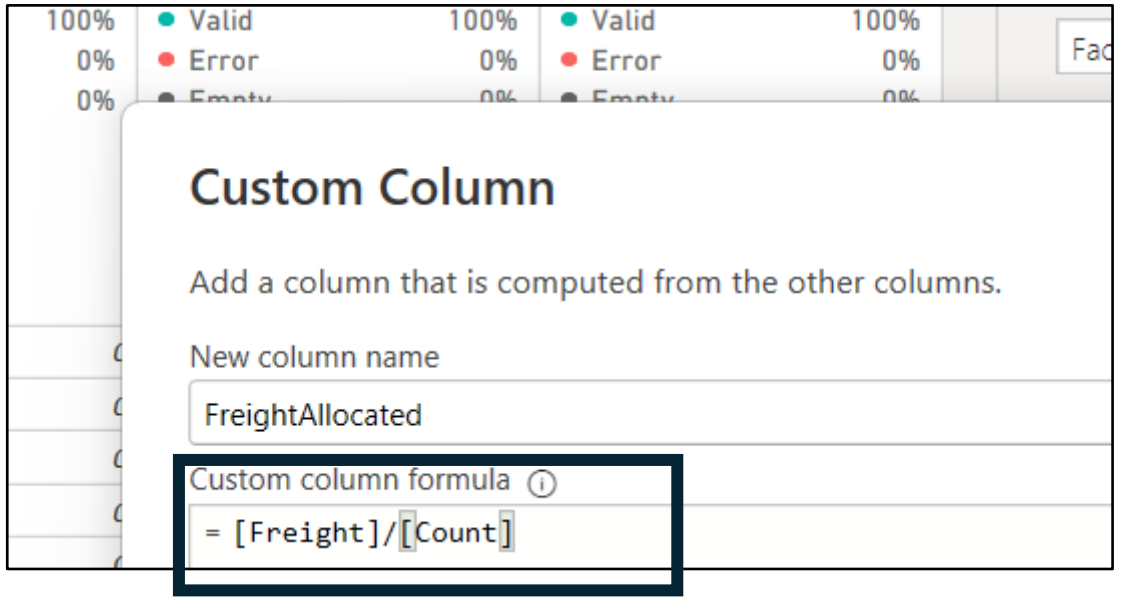
- Merge helper table with Fact table on OrderID



Allocate Header amount to the detail lines solution 1b

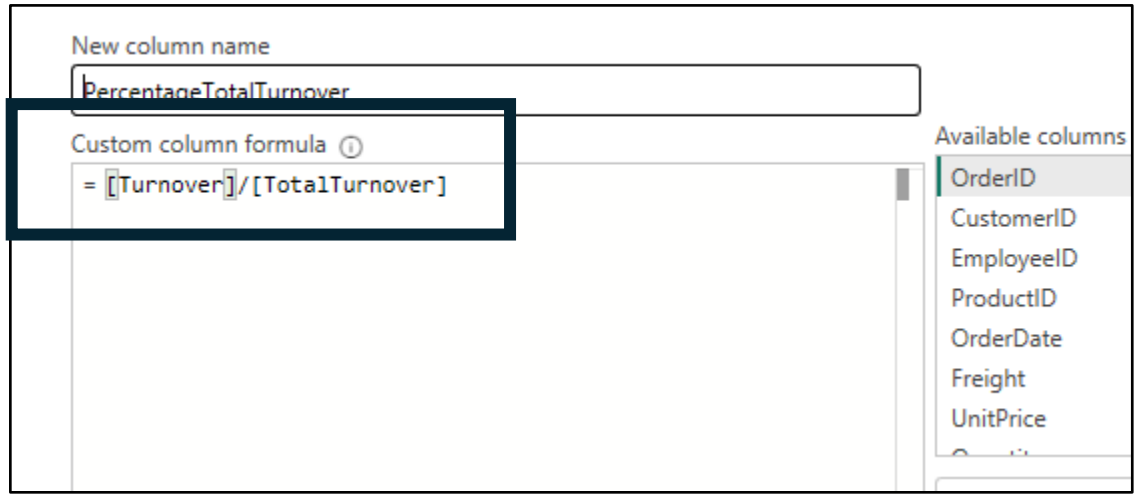
- Expand Count column
- Add custom column to divide Freight by Count
- Remove all the helper columns





Allocate Header amount to the detail lines solution 2

- Expand TotalTurnover column
- Add custom column to divide Turnover by TotalTurnover (Percentage)
- Multiply Freight with this percentage
- Remove all the helper columns



Demo time

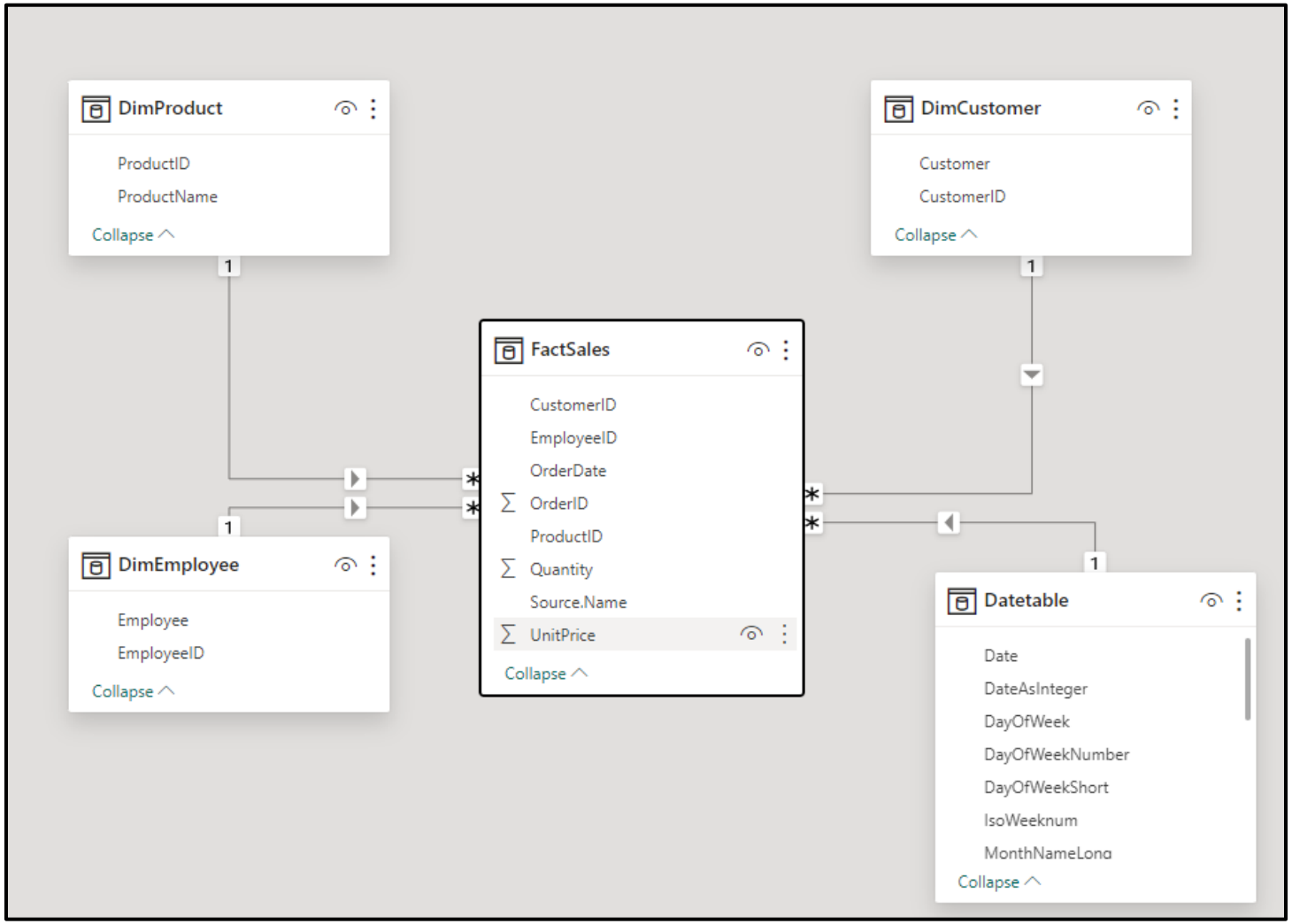
- demo_2_allocatefreight.pbix



Third demo - OneBigTable

	A	B	C	D	E	F	G
1	**CompanyName**	**LastName**	**OrderID**	**OrderDate**	**ProductName**	**Quantity**	**UnitPrice**
2	Blondesddsl père et fils	Fuller	10265	25-7-1996 00:00	Outback Lager	20	12
3	Blondesddsl père et fils	Fuller	10265	25-7-1996 00:00	Alice Mutton	30	31.2
4	Centro comercial Moctezuma	Peacock	10259	18-7-1996 00:00	Gravad lax	1	20.8
5	Centro comercial Moctezuma	Peacock	10259	18-7-1996 00:00	Sir Rodney's Scones	10	8
6	Chop-suey Chinese	Buchanan	10254	11-7-1996 00:00	Guaraná Fantástica	15	3.6
7	Chop-suey Chinese	Buchanan	10254	11-7-1996 00:00	Longlife Tofu	21	8
8	Chop-suey Chinese	Buchanan	10254	11-7-1996 00:00	Pâté chinois	21	19.2
9	Ernst Handel	Davolio	10258	17-7-1996 00:00	Mascarpone Fabioli	6	25.6
10	Ernst Handel	Davolio	10258	17-7-1996 00:00	Chang	50	15.2
11	Ernst Handel	Davolio	10258	17-7-1996 00:00	Chef Anton's Gumbo Mix	65	17
12	Ernst Handel	Dodsworth	10263	23-7-1996 00:00	Guaraná Fantástica	28	3.6
13	Ernst Handel	Dodsworth	10263	23-7-1996 00:00	Longlife Tofu	36	8
14	Ernst Handel	Dodsworth	10263	23-7-1996 00:00	Pavlova	60	13.9
15	Ernst Handel	Dodsworth	10263	23-7-1996 00:00	Nord-Ost Matjeshering	60	20.7
16	Folk och fä HB	Suyama	10264	24-7-1996 00:00	Jack's New England Clam Chowder	25	7.7
17	Folk och fä HB	Suyama	10264	24-7-1996 00:00	Chang	35	15.2
18	Frankenversand	Peacock	10267	29-7-1996 00:00	Lakkalikööri	15	14.4
19	Frankenversand	Peacock	10267	29-7-1996 00:00	Boston Crab Meat	50	14.7

The resulting Star



Demo time

- Demo_3_folder_append



Many-to-Many

- Multi-facts
- Amounts ordered & amounts paid
- Items ordered versus items delivered

Example: Many-to-Many

	A	B	C	D	E	F	G	H
1	**FullDateLabel**	**Customer**	**BrandName**	**ProductSubcategoryName**	**ProductCategoryName**	**SalesQu…ity**	**SalesAmount**	**TotalCost**
2	2007-03-31	Adventure Works	Adventure Works	Coffee Machines	Home Appliances	65	14332,268	7651,84
3	2011-11-01	BigCompany Gmbh	Adventure Works	Tobacco Machines	Home Appliances	1	100	50
4	2008-10-22	Contoso, Ltd	Contoso	Cell phones Accessories	Cell phones	2040	23504,88	12648,94
5	2009-01-31	Adventure Works	Adventure Works	Televisions	TV and Vi…	194	51593,106	28146,4

	A	B	C	E
1	**FullDateLabel**	**Customer**	**AmountPaid**	
2	2011-11-01	BigCompany	100	
3	2007-03-31	Adventure Works	14000,268	
4	2008-10-22	Contoso, Ltd	23000,88	
5	2009-01-31	Adventure Works	51593,106	

Demo time

- demo_4_multfact_bridge.pbix



Take away

- Power BI is not as easy as some people want you to believe ;-)

- Power BI is a very nice tool to **visualize**, but you need Power Query to **prepare** your underlying datamodel.
  - The more you think & act in advance, the easier the reporting will become

Reference

- https://learn.microsoft.com/en-us/power-bi/guidance/star-schema
- Youtube: https://www.youtube.com/results?search_query=star+schema+power+bi

Please evaluate..

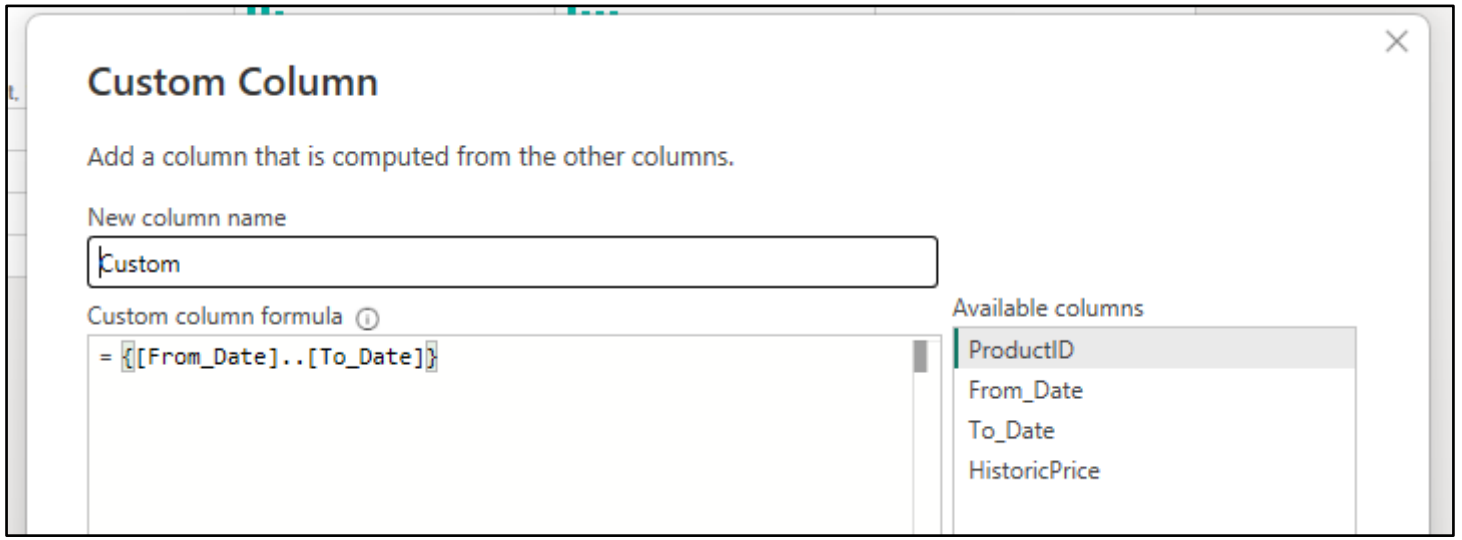


Slowly Changing Dimension

	A	B	C	D
1	**ProductID**	**From_Date**	**To_Date**	**HistoricPrice**
2	2	1-7-1996	21-7-1997	19,00
3	2	22-7-1997	31-12-1997	20,50
4	2	1-1-1998	31-7-2024	23,00
5	5	1-7-1996	31-12-1997	21,00
6	5	1-1-1998	31-7-2024	22,50

Slowly Changing Dimension

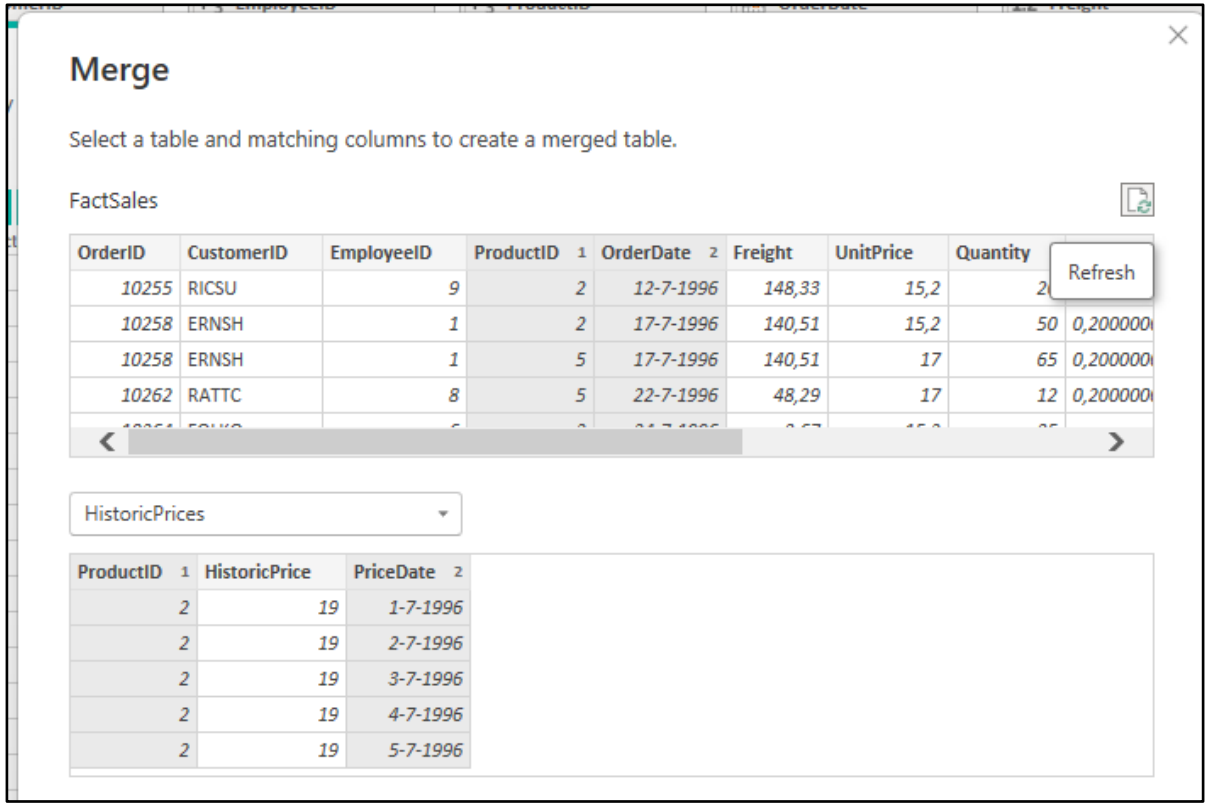
- Transform dates between From_Date and To_Date into Rows
- **`{[From_Date]..[To_Date]}`**



20	2	19	20-7-1996
21	2	19	21-7-1996
22	2	20,5	22-7-1996
23	2	20,5	23-7-1996

Slowly Changing Dimension

- Merge Historic Price into Fact table on ProductID and Date



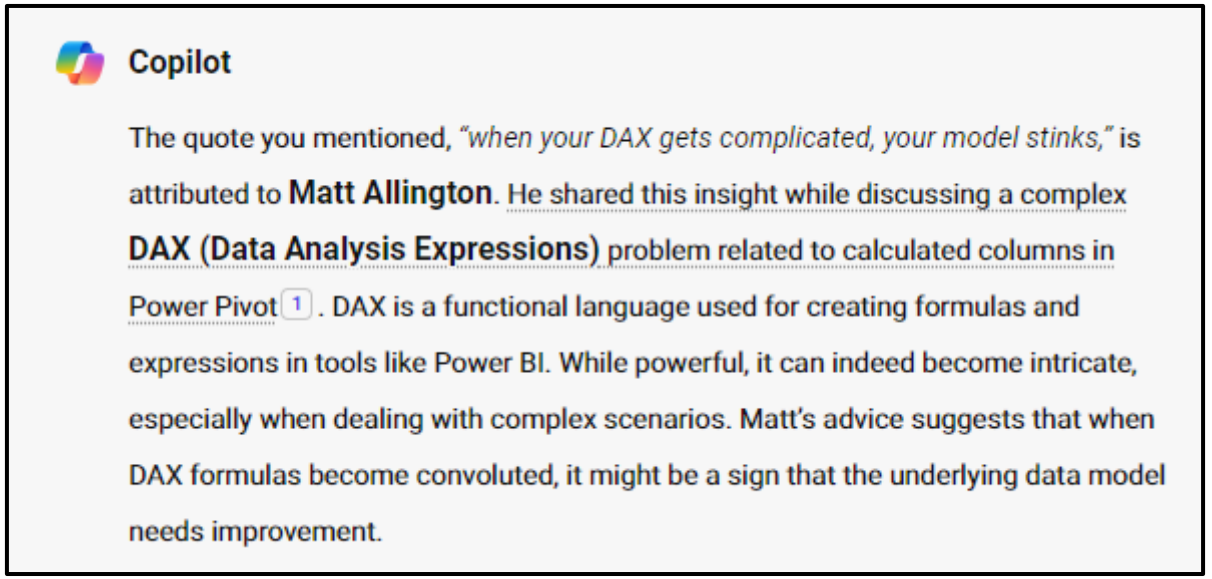
OrderID	CustomerID	EmployeeID	ProductID 1	OrderDate 2	Freight	UnitPrice	Quantity
10255	RICSU	9	2	12-7-1996	148,33	15,2	2
10258	ERNSH	1	2	17-7-1996	140,51	15,2	50
10258	ERNSH	1	5	17-7-1996	140,51	17	65
10262	RATTC	8	5	22-7-1996	48,29	17	12

ProductID 1	HistoricPrice	PriceDate 2
2	19	1-7-1996
2	19	2-7-1996
2	19	3-7-1996
2	19	4-7-1996
2	19	5-7-1996

Demo time



Summary: “when your DAX gets complicated, your model stinks”



Copilot

The quote you mentioned, *“when your DAX gets complicated, your model stinks,”* is attributed to **Matt Allington**. He shared this insight while discussing a complex **DAX (Data Analysis Expressions)** problem related to calculated columns in Power Pivot [1]. DAX is a functional language used for creating formulas and expressions in tools like Power BI. While powerful, it can indeed become intricate, especially when dealing with complex scenarios. Matt’s advice suggests that when DAX formulas become convoluted, it might be a sign that the underlying data model needs improvement.



Marco Russo • 1e

Business Intelligence trainer and consultant

Italy

40K volgers



113 gemeenschappelijke connecties

Bericht | Volledig profiel weergeven

Bijdragen van Marco



Marco Russo • 1e

Business Intelligence trainer and consultant

6 u

Short, sad story.

Once upon a time, there was a ***lazy*** data modeler.
Their semantic model did not work.

End of the story.

Credits: Alberto Ferrari

#powerbi #datamodeling

U en 251 anderen — 11 commentaren • 7 reposts

Interessant | Commentaar | Reposten | Versturen