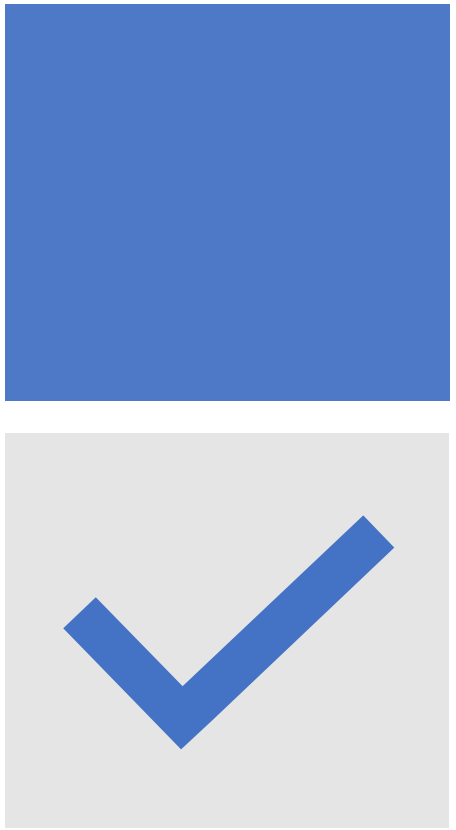
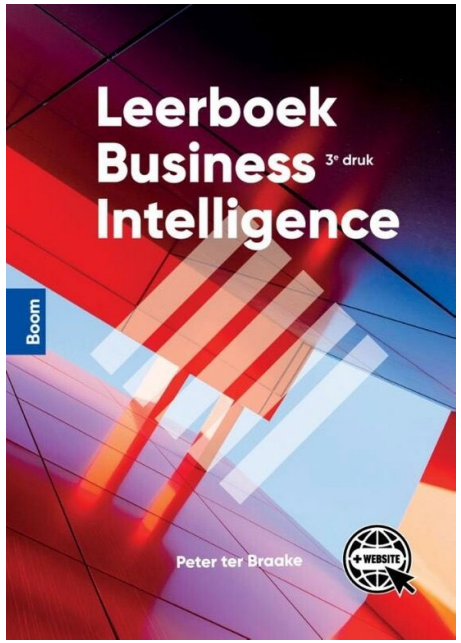
Data modeling: The key to success with Power BI

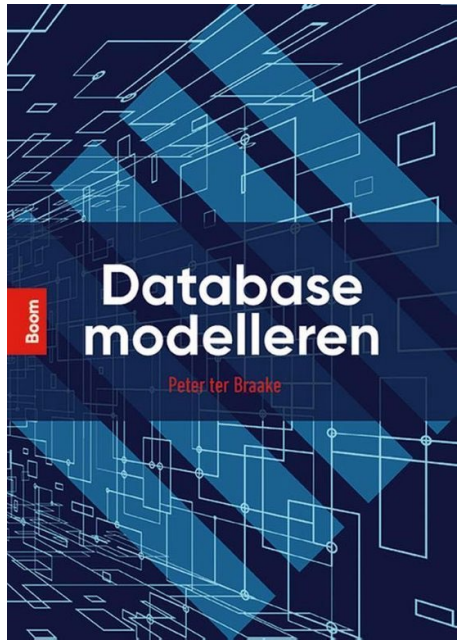


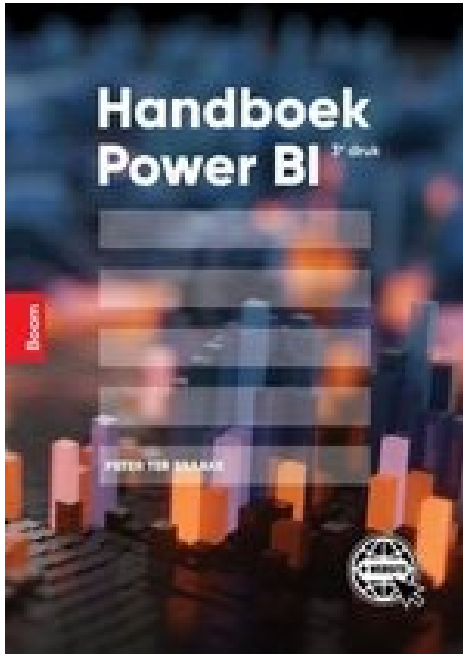
Peter ter Braake

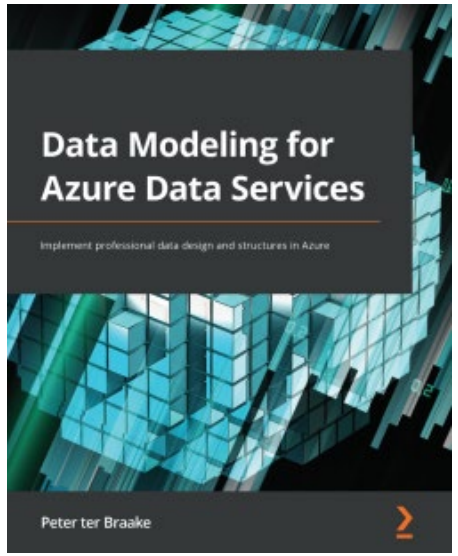
Introducing myself

- Peter ter Braake
  - trainsql@live.nl
  - Independent contractor
  - MCT since 2002
- Microsoft Data Platform
  - SQL Server
    - Incl. SSRS
  - Business Intelligence
  - Power BI
- Author









Quote

Alberto Ferrari in SQLBI.com:

The golden rule of data modeling is always the same: always use star schemas.

- If a column has to be used to slice and dice, then it needs to belong to a dimension.
- Numbers to aggregate, on the other hand, are stored in fact tables.

Tabular lets a developer deviate from the regular star schema architecture. This does not mean that doing it is always a good idea. It seldom is.

Power BI scenarios

- Corporate Reporting
  - Managers use reports created by BI profesisonals
- Self service Collaboration
  - Collaborate on shared dataset
- …
- Self service analist / data scientist
  - Friday afternoon 4 O'Clock …

Agenda

- Theory
- Discussion

Why do we need a DWH?

Performance

- For both OLTP and OLAP

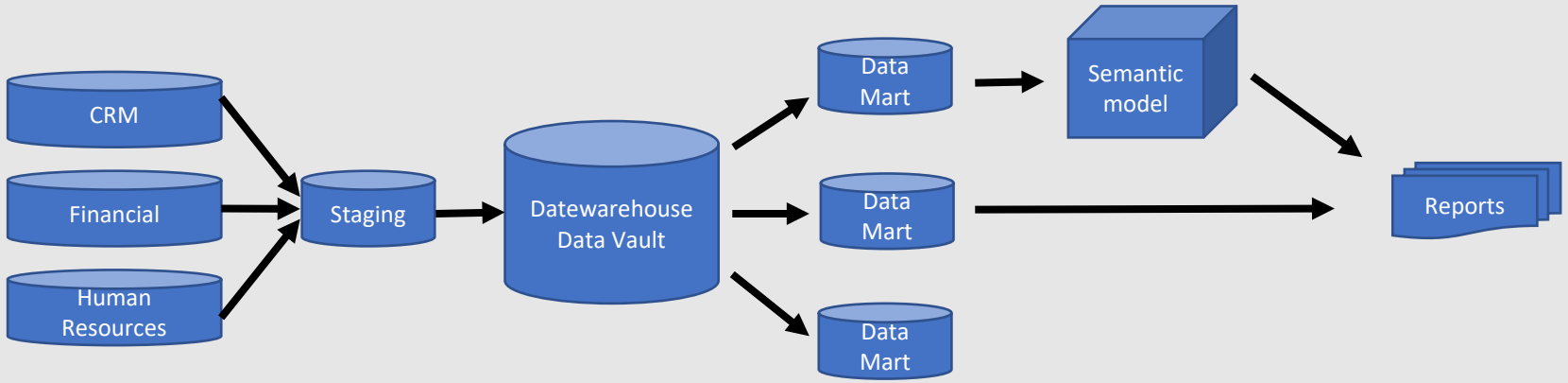
Consistency across systems (stove pipes)

Data Quality

Data model complexity

Historical data

Power BI versus Data warehouse



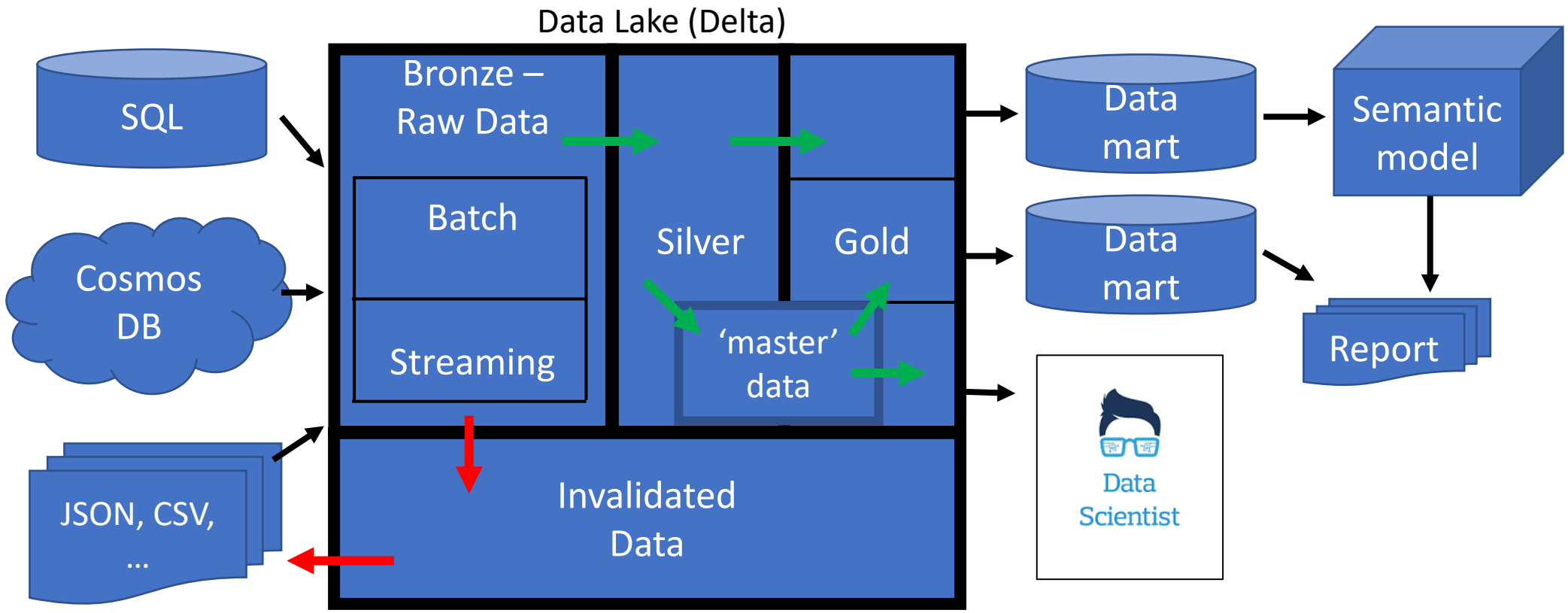
Source → ETL → BISM → Report

Power BI:

Source → Get Data → Power Query → DAX → Report

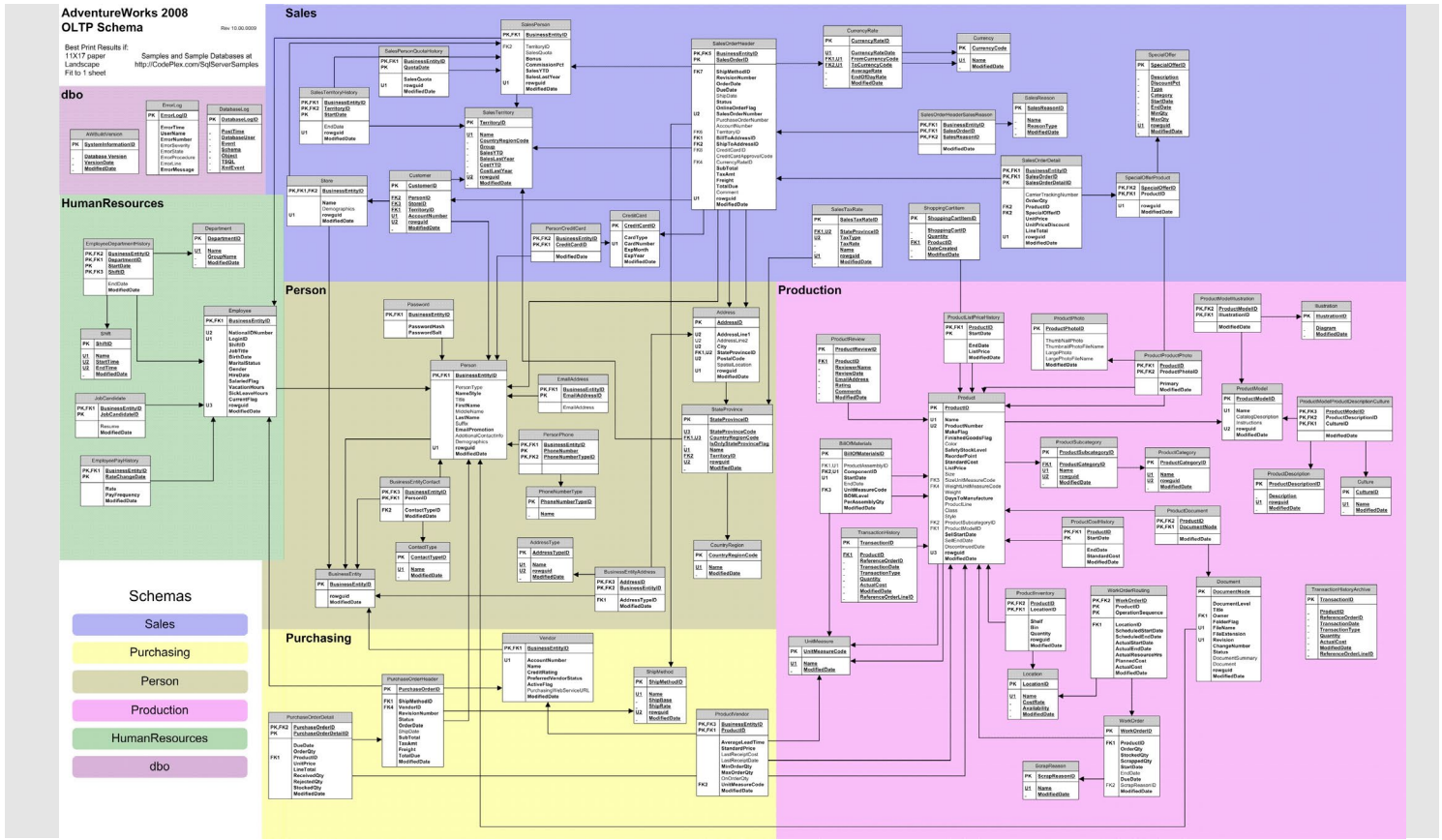
- Same architecture, same best practices
- Implement solutions as far to the left as possible

Architecture Overview



Dimensional modelling

Why normalize data?



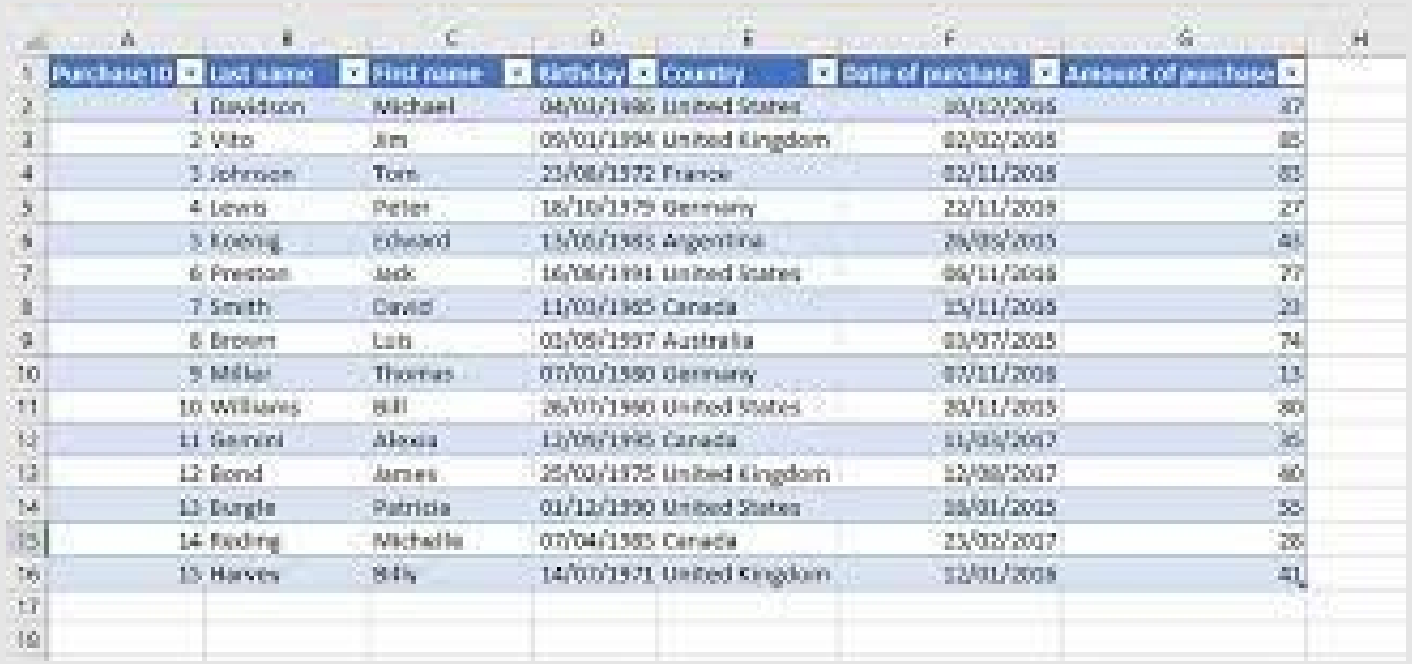
Normalizing is based on:

- Reduce redundancy
- Model dependencies between attributes

Data warehouse / reporting / analytics:

- Read-only workload
  - Redundancy might be beneficial
- Inefficiency in storing redundant data is small
  - Especially for Power BI / tabular / Columnstore
- Make usage (processes) important instead of putting the data central

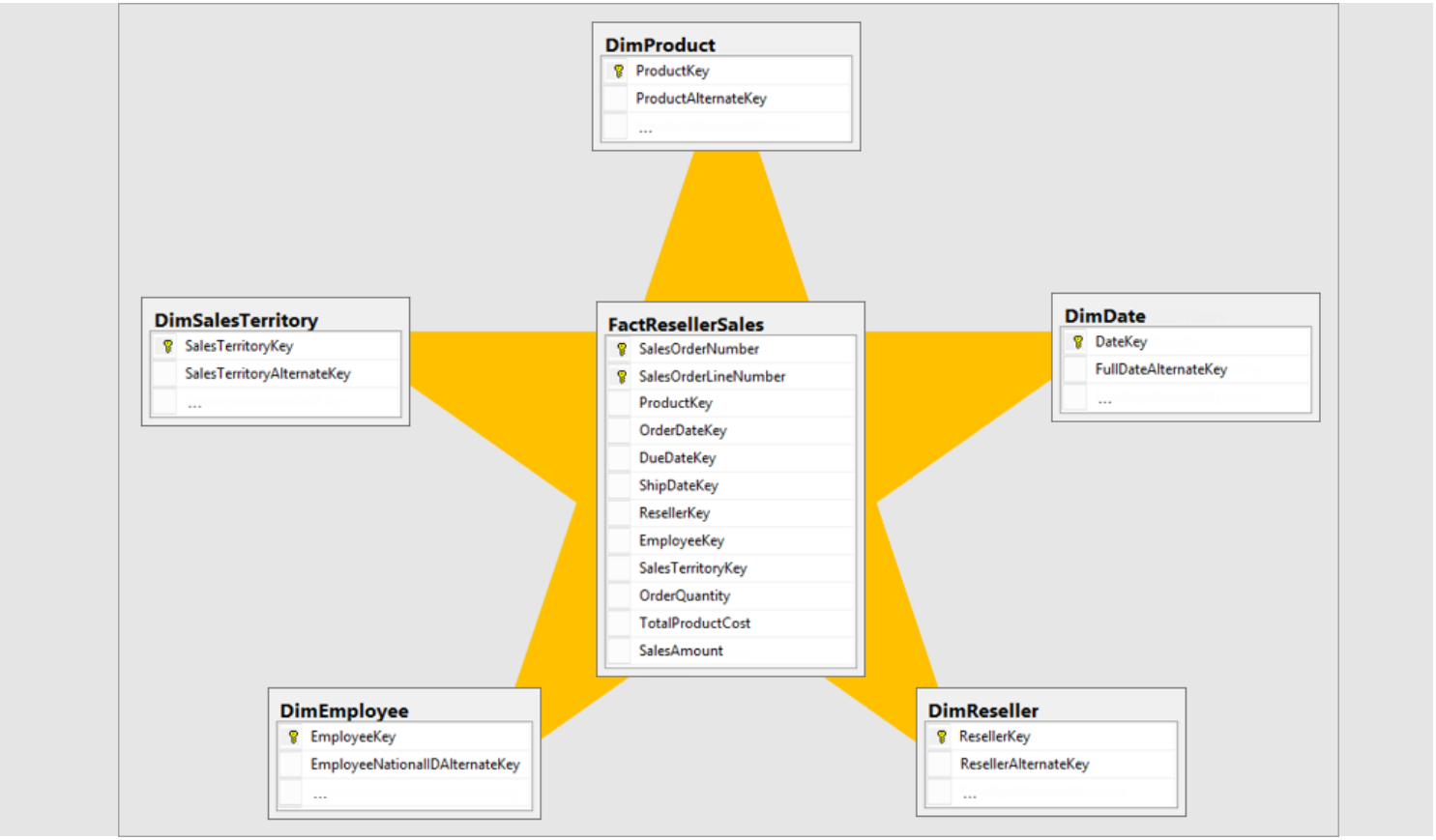
Excel?



Purchase ID	Last name	First name	Birthday	Country	Date of purchase	Amount of purchase
1	Davidson	Michael	04/01/1980	United States	10/12/2005	37
2	Vito	Jim	05/01/1994	United Kingdom	02/02/2005	85
3	Johnson	Tom	23/08/1972	France	02/11/2005	65
4	Lewis	Peter	18/10/1975	Germany	22/11/2005	27
5	Koenig	Edward	13/06/1985	Argentina	26/05/2005	41
6	Preston	Jack	16/08/1991	United States	06/11/2005	77
7	Smith	David	11/03/1985	Canada	15/11/2005	23
8	Brown	Lish	01/05/1957	Australia	03/07/2005	74
9	Miller	Thomas	07/01/1980	Germany	07/11/2005	13
10	Williams	Bill	26/07/1960	United States	20/11/2005	89
11	Gerrini	Alexia	13/06/1995	Canada	11/05/2007	35
12	Bond	James	25/02/1975	United Kingdom	12/08/2007	80
13	Burgle	Patricia	01/12/1990	United States	28/01/2005	55
14	Fading	Michelle	07/04/1985	Canada	23/02/2007	28
15	Harves	Billy	14/07/1971	United Kingdom	12/01/2005	41

- Popularity based on Pivot table
- Pivot table based on single table containing columns used
  - On rows
  - On columns
  - On filters
  - As values
    - Using aggregation

Why star schema?



- Intuitive / simple
  - Less error prone
  - Rapid report development
- Better performance
  - Less ‘joins’
- Power BI
  - More consistent experience
  - Easier DAX
  - ‘auto exist’
  - Effect cross filter both ways
  - Better performance

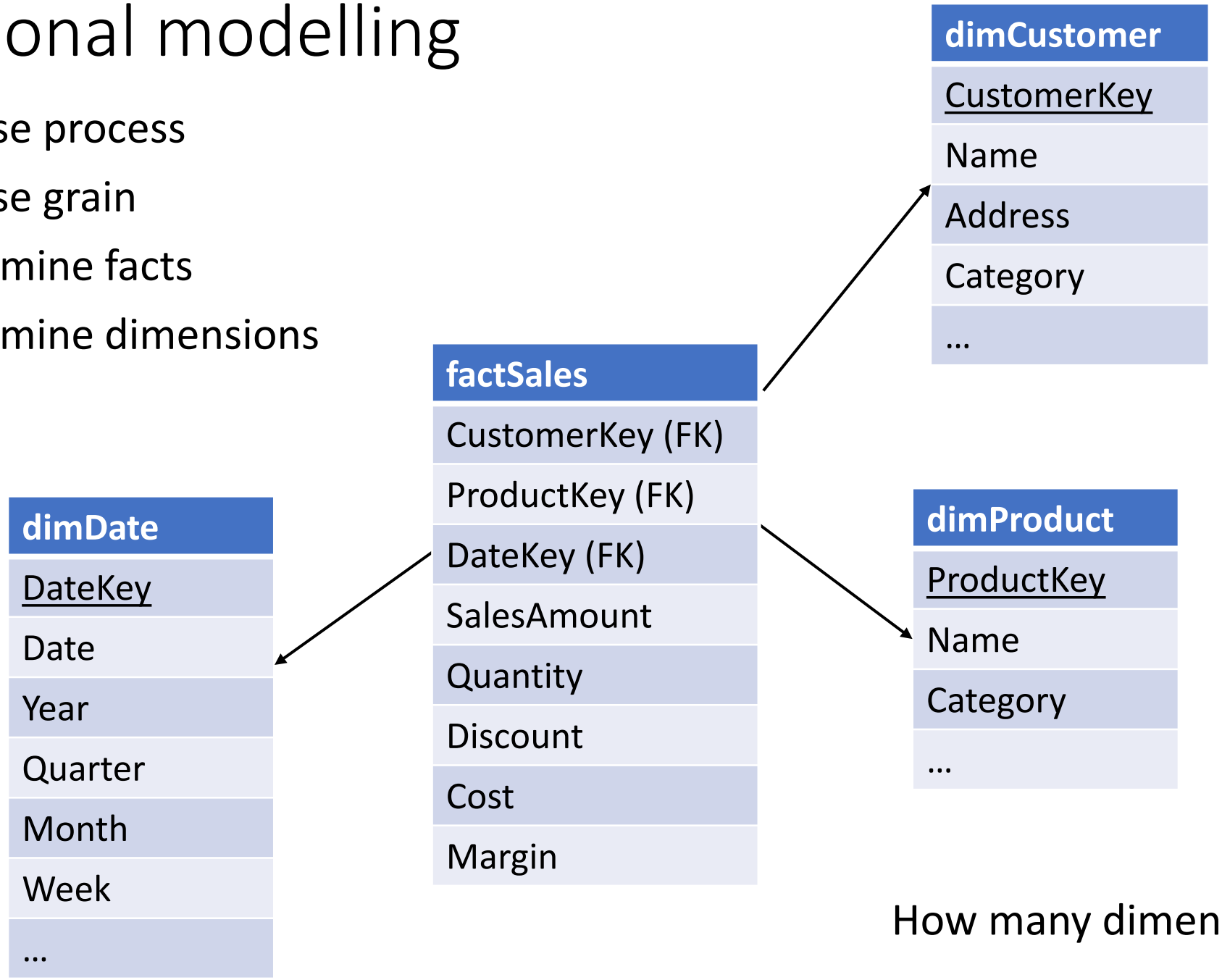
Dimensional modelling

Step 1: Choose process

Step 2: Choose grain

Step 3: Determine facts

Step 4: Determine dimensions



How many dimensions?

Step 1: Choose grain

- Level of detail
- Largely determines size of semantic model
  - OneLake storage is not cheap
  - Size of imported data limited
- Determines efficiency of aggregations
  - CUs are not cheap in Fabric
- Consider Power BI Aggregations

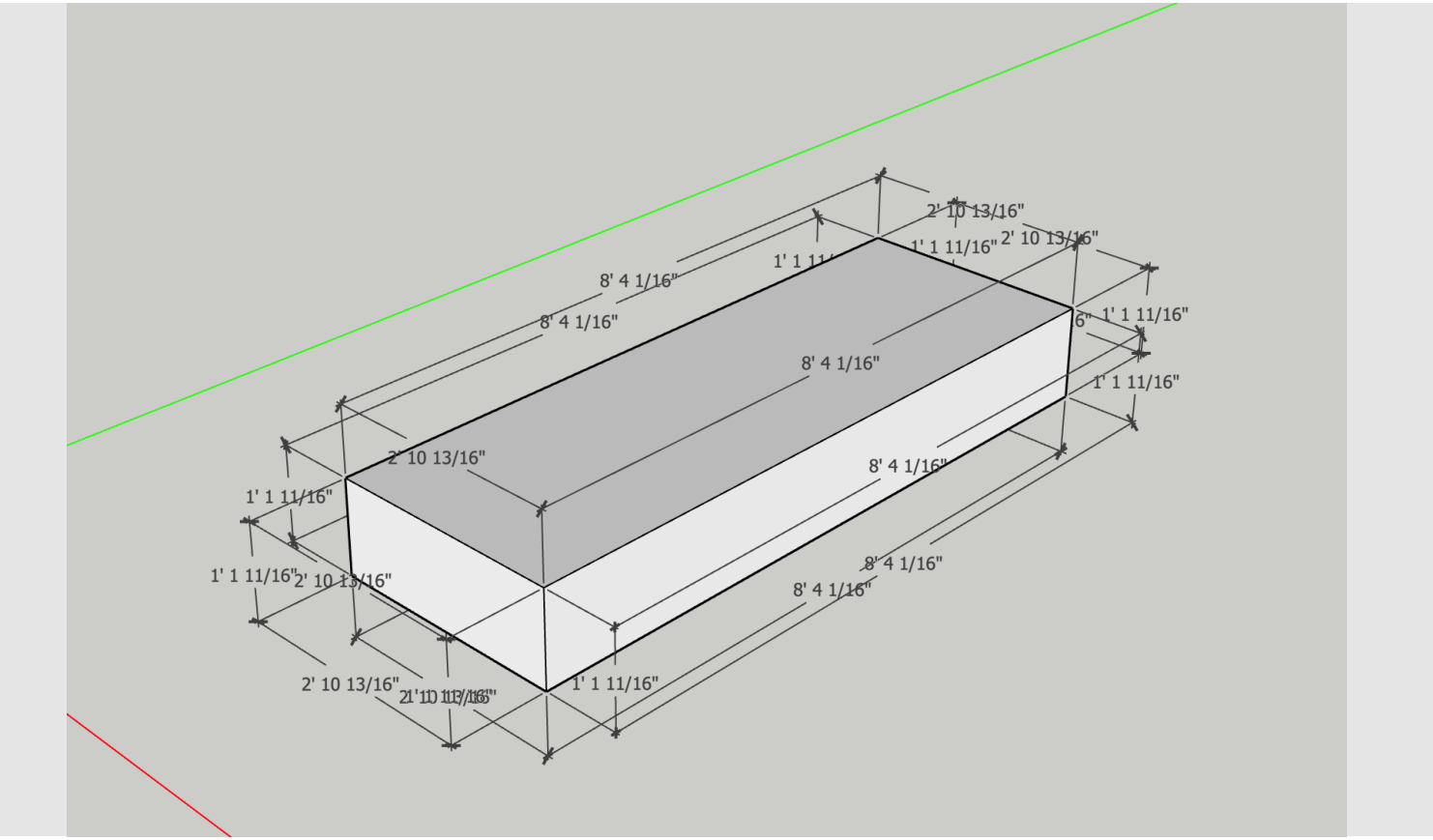
Step 2: Choose facts

- What do you analyze?
- Favor additive facts
- Favor Fixed decimal number over decimal number
- Create your own ‘smart’ facts

Demo

Centuries

Step 3: Create dimensions

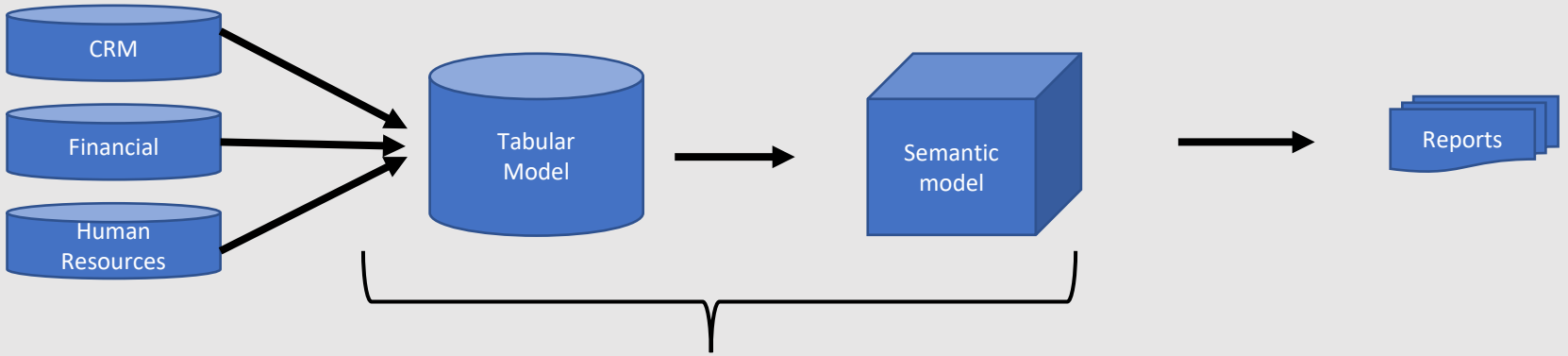


- Analytical power of model
  - Determines functionality
  - Determines interpretability
- Consider cardinality
  - Grouping / binning
  - Date / Time / DateTime
- Create efficient keys
  - Whole number / Date
- Only snowflake when necessary
- Use unknown / not applicable member
- Create proper date dimension
  - Mark as date table
- Consider SCD

Demo

Year index

Power BI: where, what!



Source	Power Query	DAX	Report
Crap!	Facts and dimensions	Enrich model	
Needs cleansing	Good in row-by-row operations	Good in aggregations	Good in visualizations

Find solutions as far left as possible

Calculated column versus Calculated measure

Feature	Calculated Column	Calculated Measure
Calculation Type	Row-by-row calculations based on each row of data	Context-sensitive, calculated based on the current report context
Storage	Stored in the data model as part of the table	Not stored; calculated on the fly during report interaction
Calculation Timing	Calculated during data refresh and stored permanently	Calculated dynamically when you interact with visuals
Scope	Can be used in filters, rows, columns, or slicers like a regular column	Primarily used for aggregations and summarizations in visuals
Performance Impact	Can increase model size due to storage of calculated values	Does not increase model size, but complex calculations can impact performance
Example Use Case	Creating a new column based on conditions (e.g., categorizing data)	Aggregating data dynamically (e.g., total sales, average)

- Calculated column
  - Row context
  - Static
  - Calculated once during data refresh
- Calculated measure
  - Filter context
  - Dynamic
  - Calculated every time needed in visual

Demo

of products

Discussion: fact or dimension

UnitPrice

Afstand

FTE

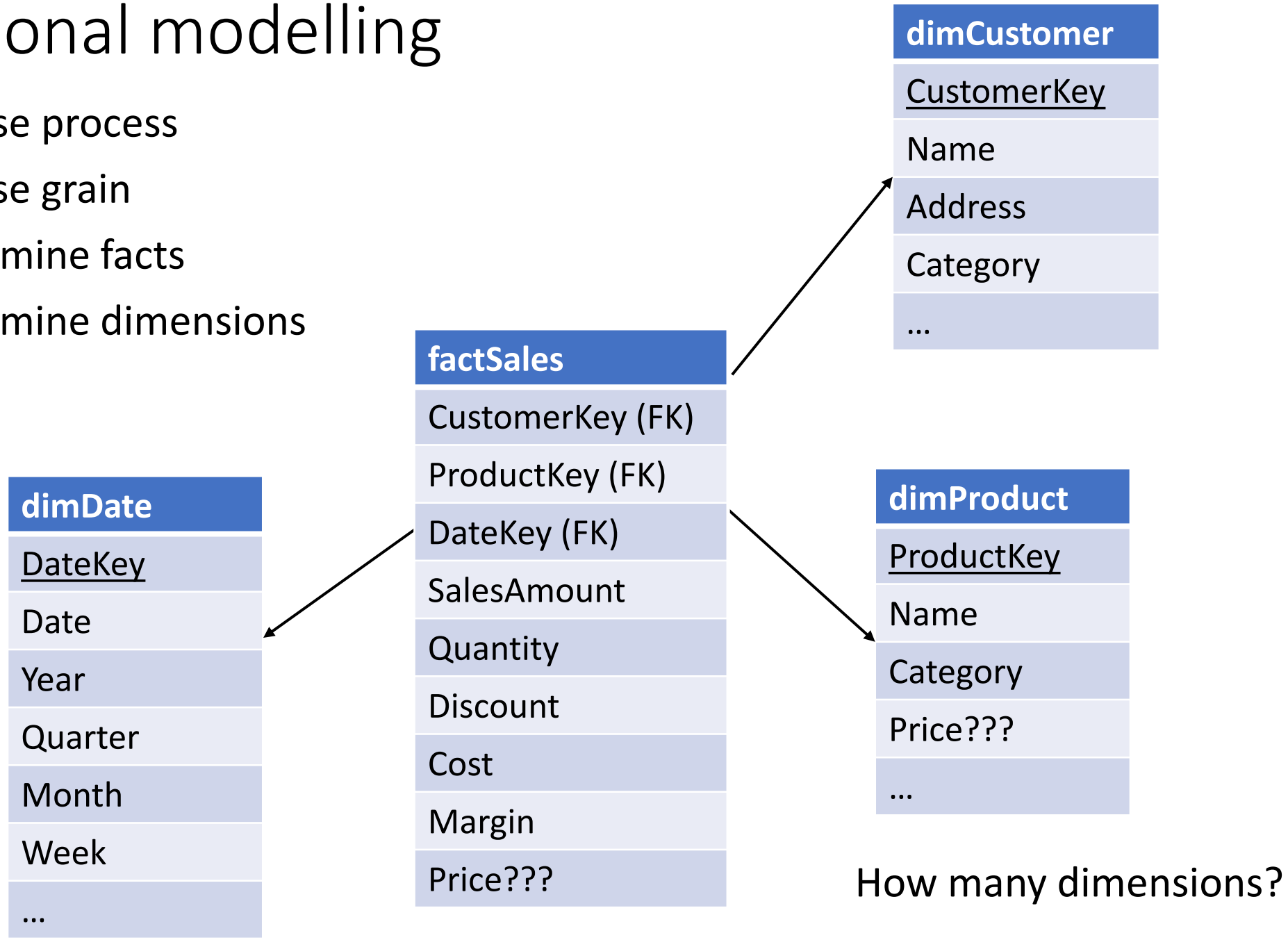
Dimensional modelling

Step 1: Choose process

Step 2: Choose grain

Step 3: Determine facts

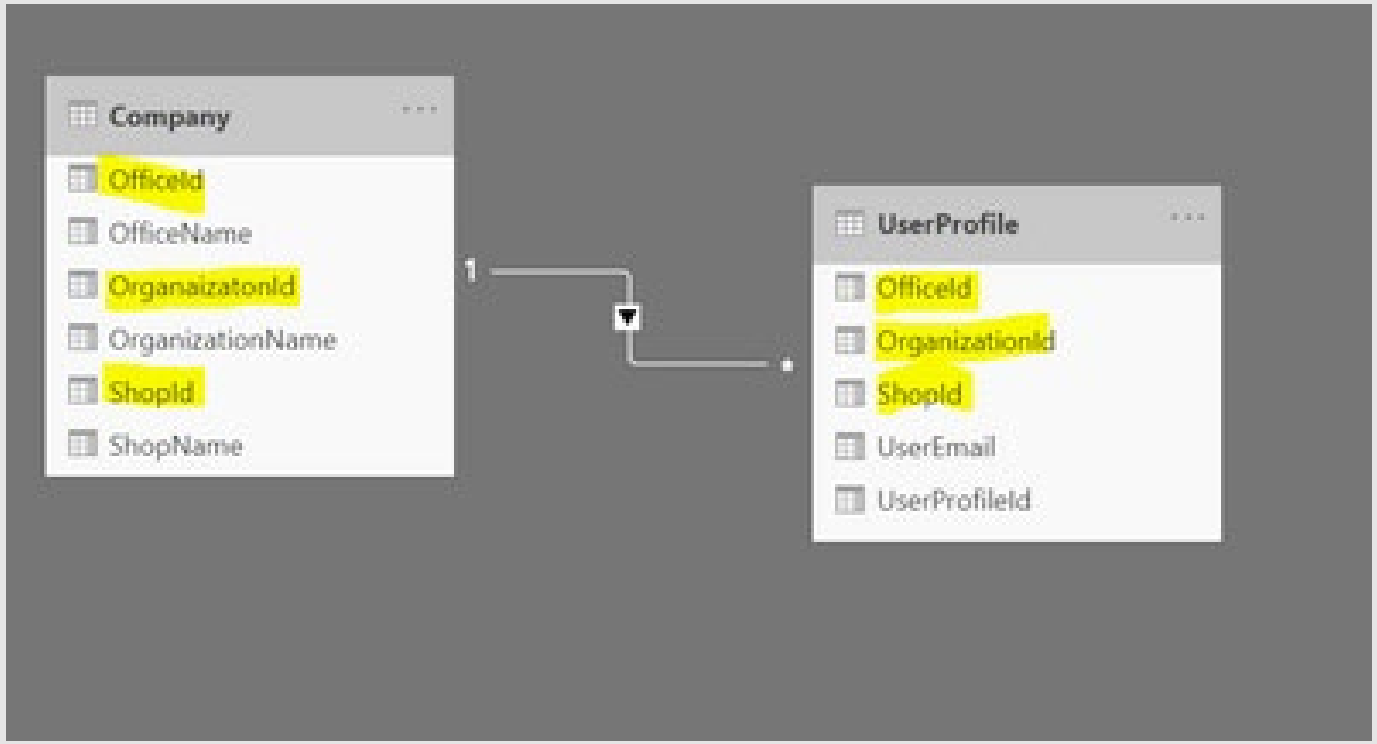
Step 4: Determine dimensions



How many dimensions?

Relationships

Multi-column relationship



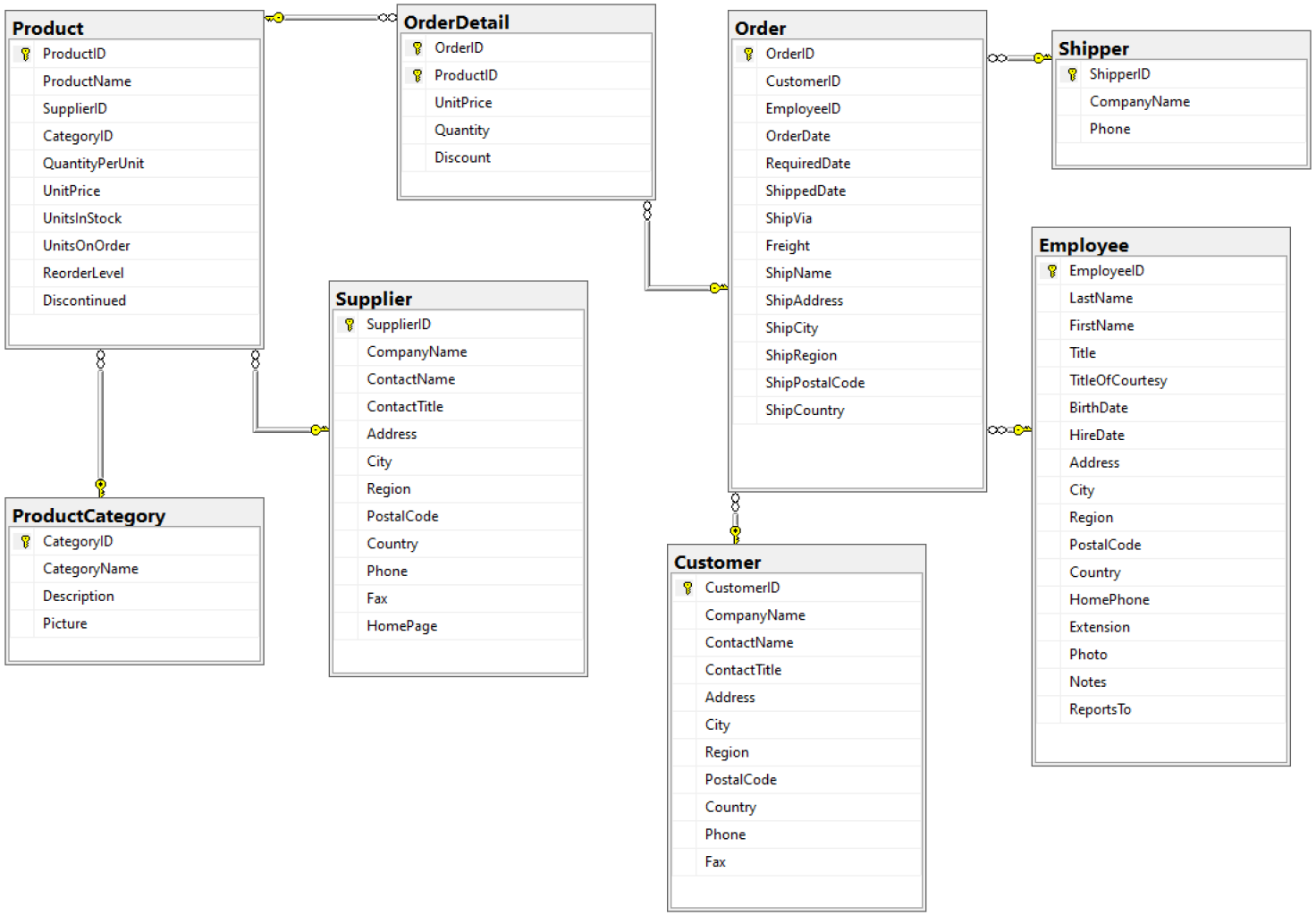
- Import:
  - Use PowerQuery (ETL tool) to create single column
- DirectQuery
  - Use COMBINEVALUES to create single column
  - ```
    Table1[CalcColumn]
    =
    COMBINEVALUES(",",
    Table1[Column1],
    Table1[Column2])
    ```
    ```

Relationship type

- Automatically inferred
- Regular relationships
  - Query engine can determine the "one" side
  - All one-to-many intra source group relationships
- Limited relationships
  - No guaranteed "one" side
    - Many-to-many
    - Cross source group
      - Composite models

Discussion (if time permits)

Northwind



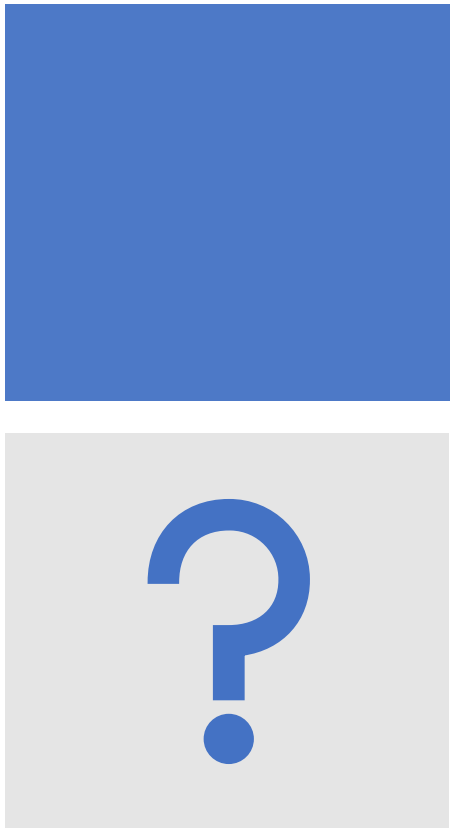
- Choose process
- Define grain
- Determine facts
- Determine dimensions

Wrap-up

Sum-up

- Use a star schema. Always.
- Hide all the columns in the fact table. Only show measures in the fact table.
- Expose visible attributes only through dimensions.
- Test your formulas on small sets of data that you can master and understand at a glance.

Questions ?



Thank you

Peter ter Braake

trainsql@live.nl

Power BI en de wereld van Excel als bron

valcon

Geanpaul Mercera

Noa van Hunen

valcon

JungleBoots Factory heeft dringend behoefte aan inzicht in de verkoopgegevens van haar groeiende aantal winkels om het "Get your jungleboots" programma effectief te kunnen sturen. Momenteel ontbreekt het bedrijf aan een database of rapportagesysteem om de verkoop van berg- en wandelschoenen per winkel te analyseren.

Het management heeft echter zo snel mogelijk inzicht nodig in de verkoopcijfers om de juiste beslissingen te kunnen nemen. Hoe gaan wij JungleBoots Factory helpen om een oplossing te vinden waarmee ze snel en efficiënt verkooprapportages kunnen genereren, ondanks het ontbreken van een database?


valcon

Hoe gaan we dit doen?

valcon

Future proof maken

- Consolideren van data
- Data verrijking
- Data verversing waarborgen (datakwaliteit)
- Business proof

valcon

valcon

valcon


Inladen met M (Power Query)
valcon

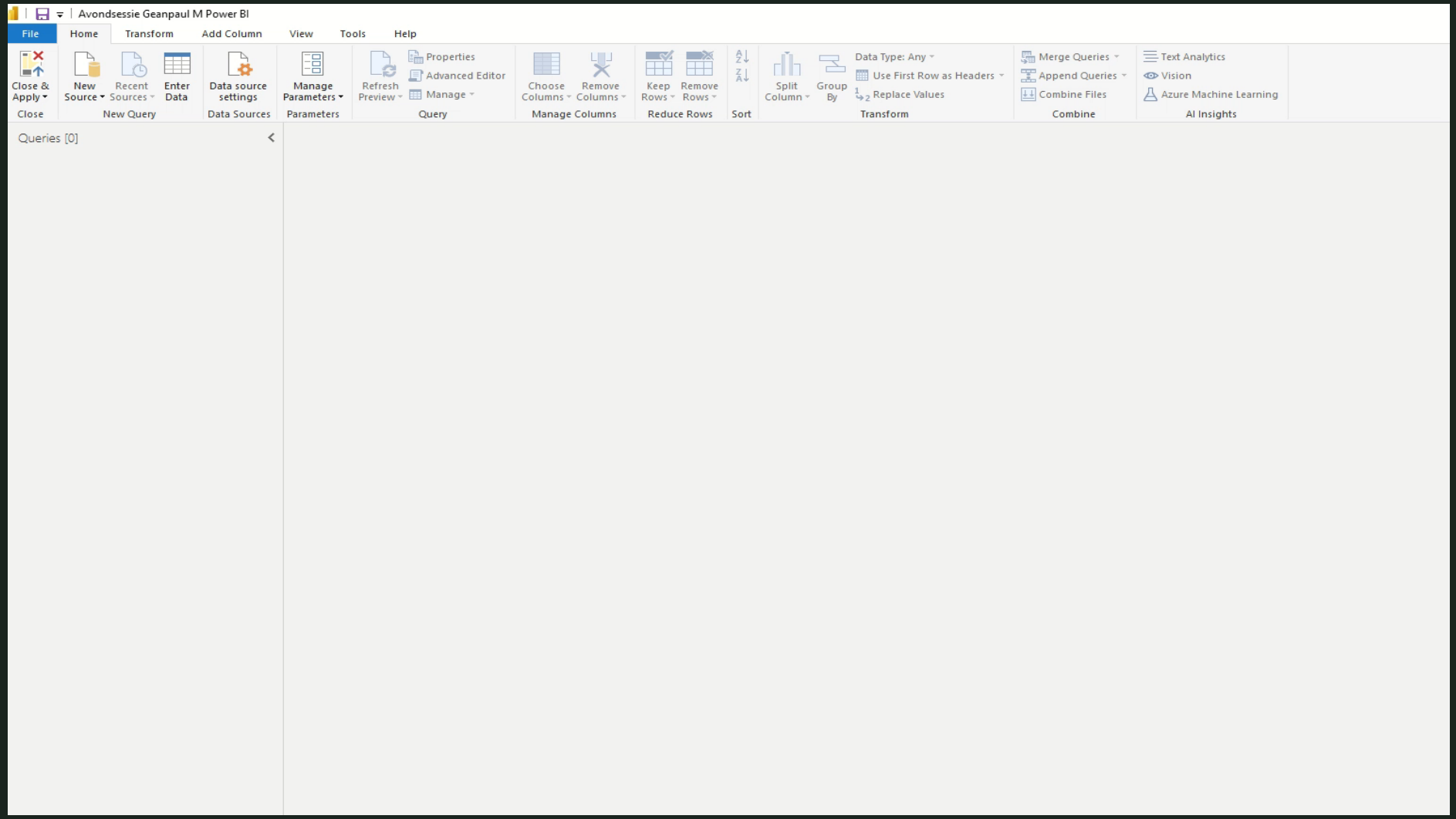
Data opslag

valcon


Documenten > General > Verkoop Data Winkels
Naam
Gewijzigd
Gewijzigd door
Amsterdam.xlsx
3 days ago
Geanpaul Mercera
Den Haag.xlsx
3 days ago
Geanpaul Mercera
Utrecht.xlsx
3 days ago
Geanpaul Mercera

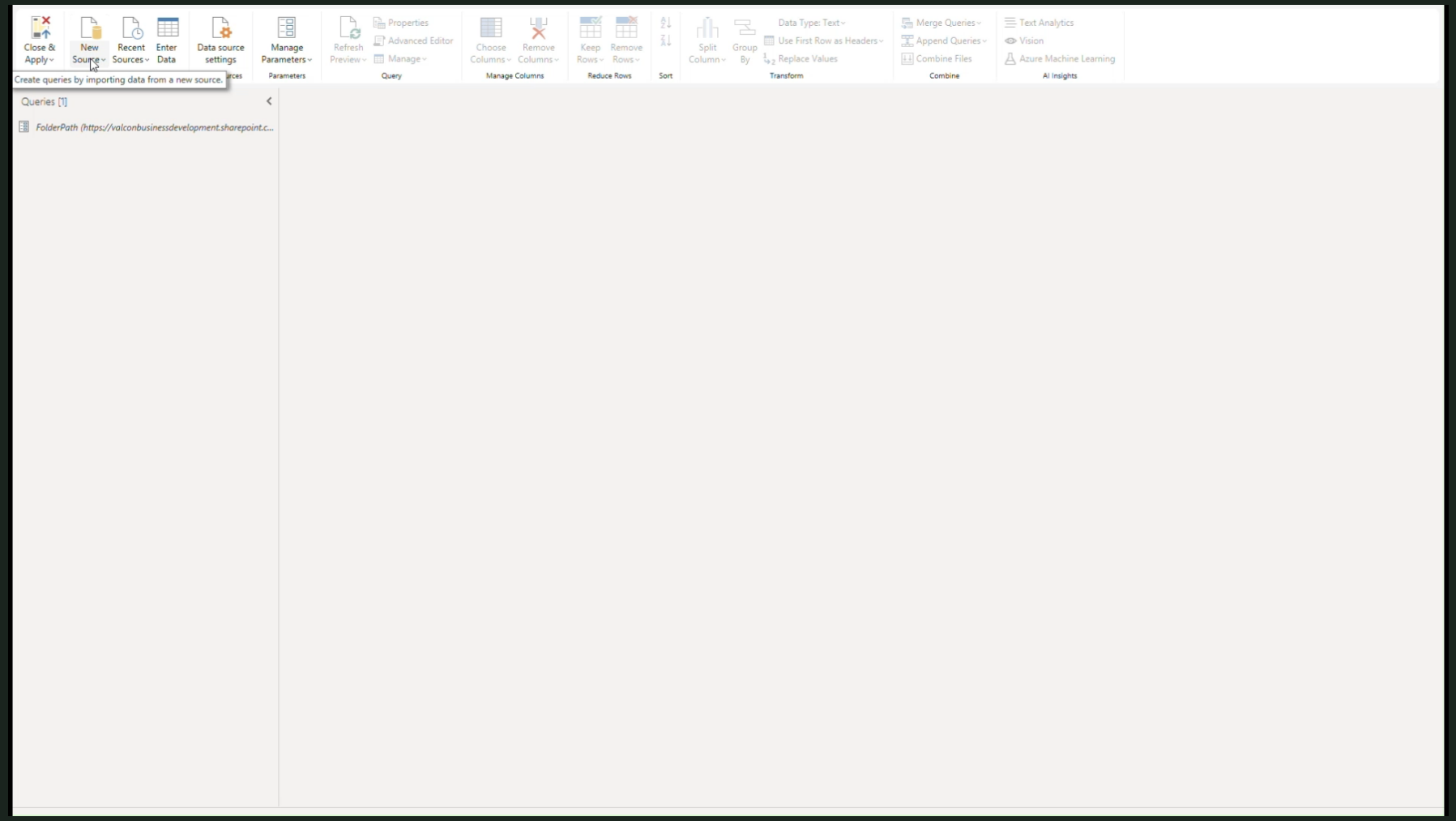
valcon

Parameters

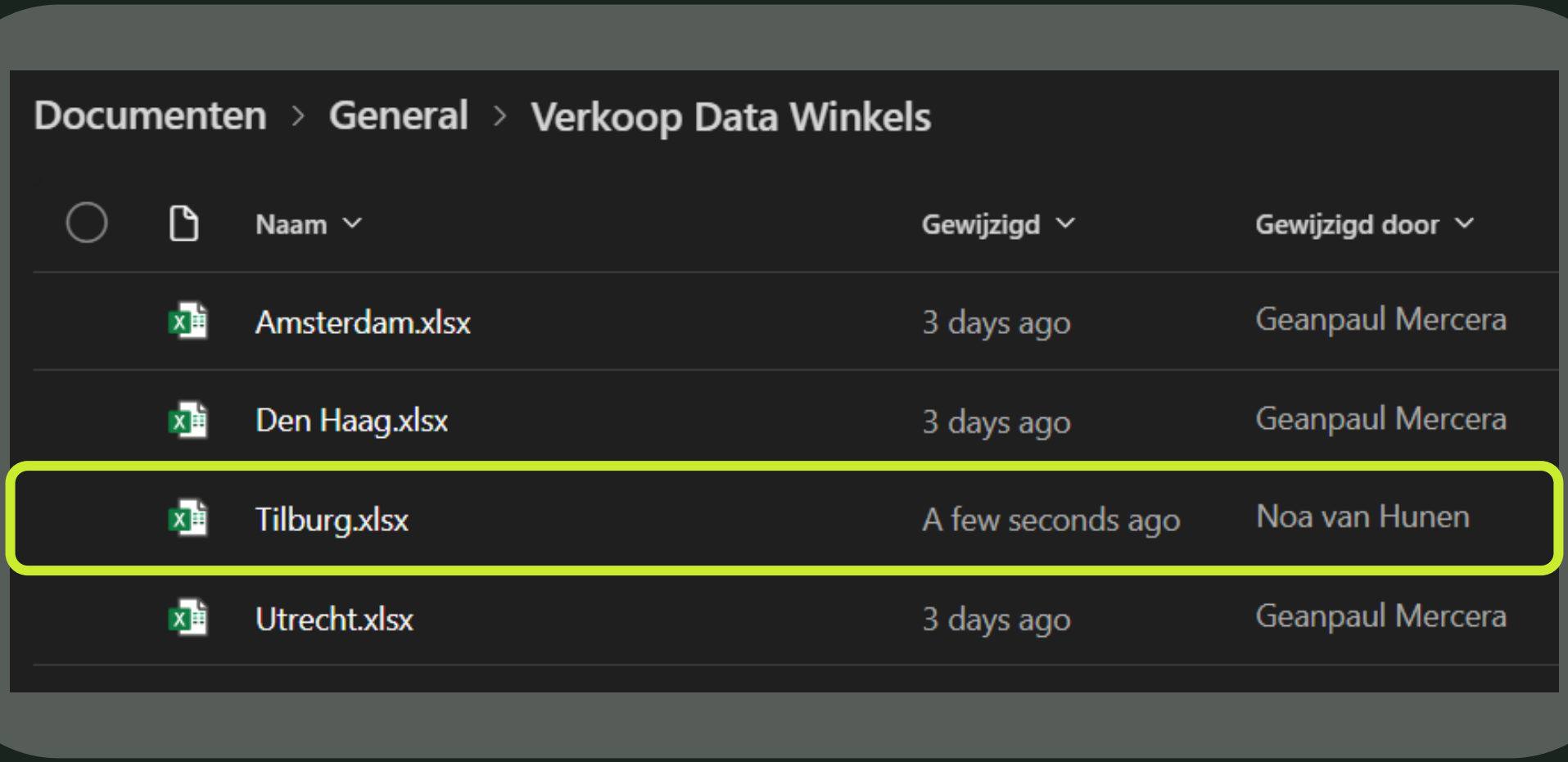


valcon

Inladen

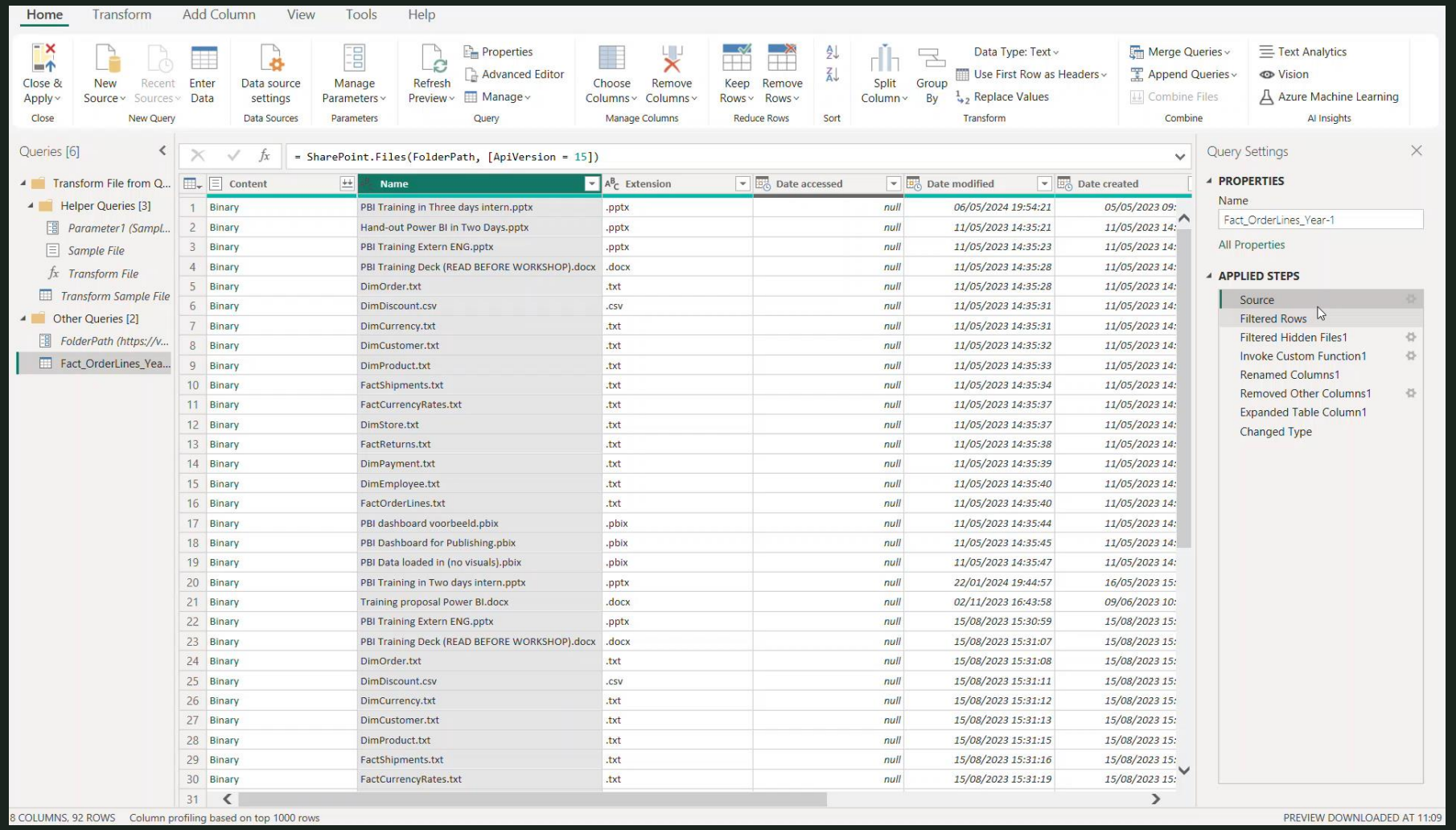


valcon


Documenten
General
Verkoop Data Winkels
Naam
Gewijzigd
Gewijzigd door
Amsterdam.xlsx
3 days ago
Geanpaul Mercera
Den Haag.xlsx
3 days ago
Geanpaul Mercera
Tilburg.xlsx
A few seconds ago
Noa van Hunen
Utrecht.xlsx
3 days ago
Geanpaul Mercera

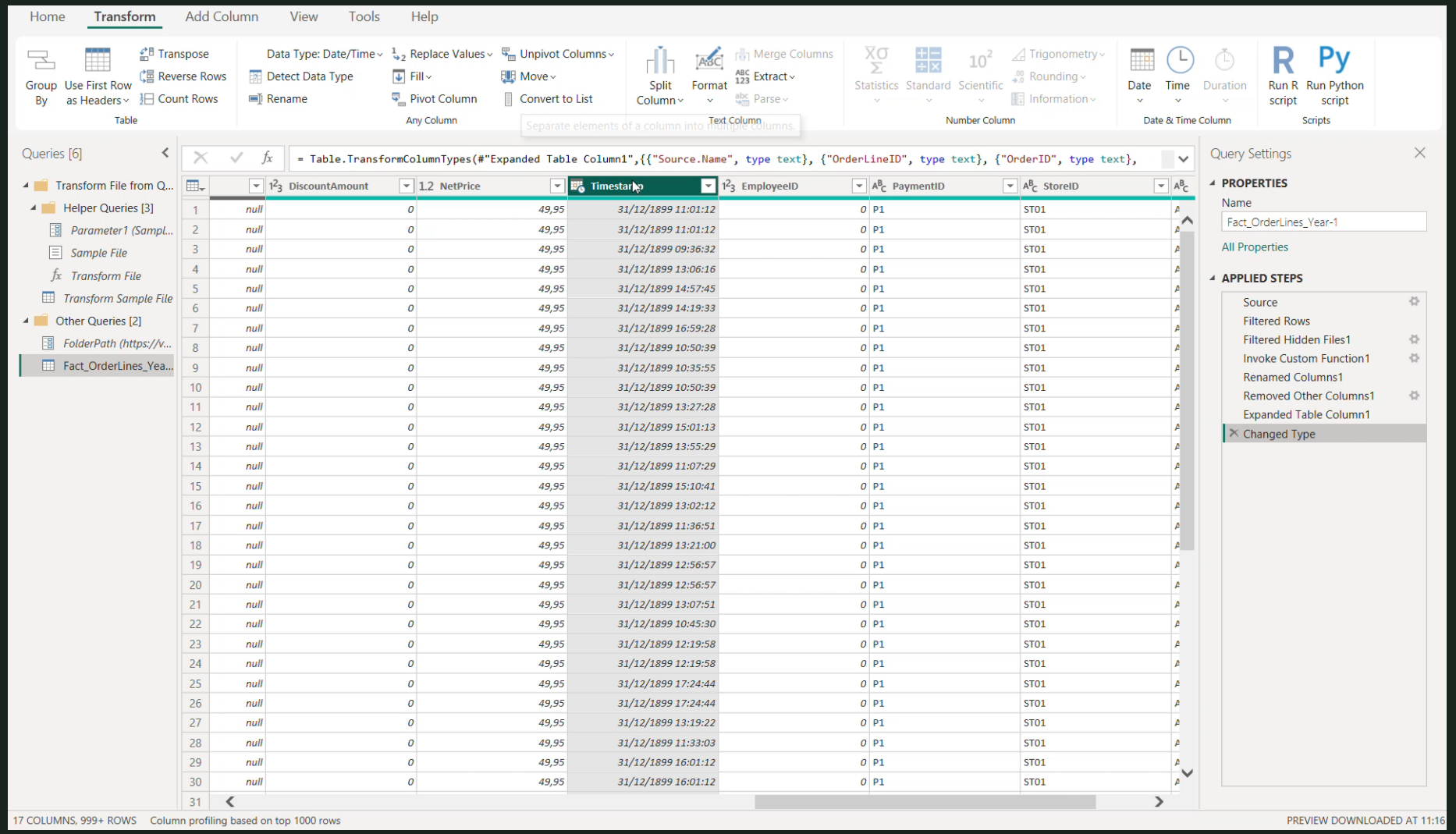
valcon

Nieuwe winkel: nieuw bestand




Data Cleansing met M
valcon

Split en trim



valcon

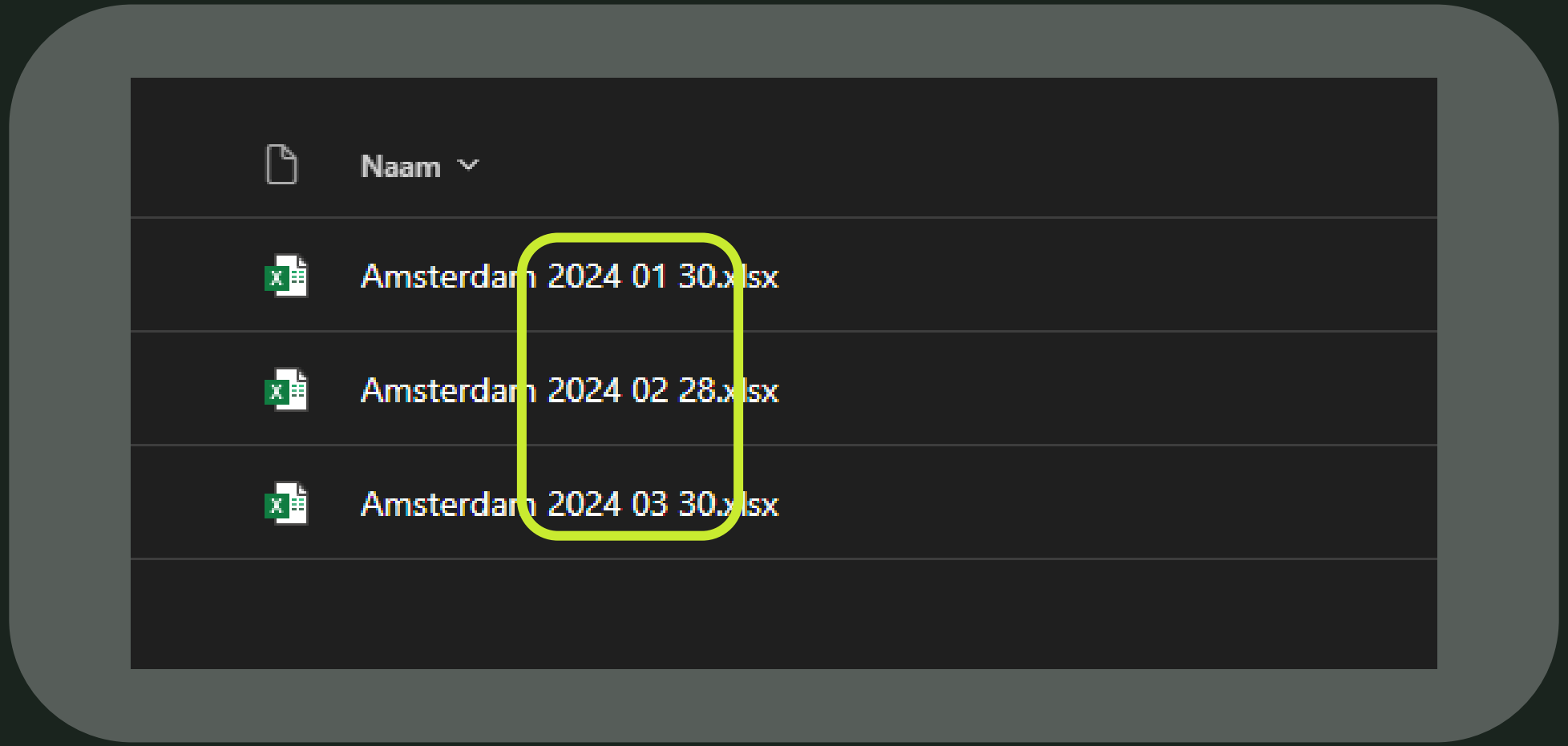
Veel gebruikte M functies

Text.Contains
Text.End
Text.Length
Text.Replace
Text.Trim
Text.Start
Date.From
Date.FromText

valcon

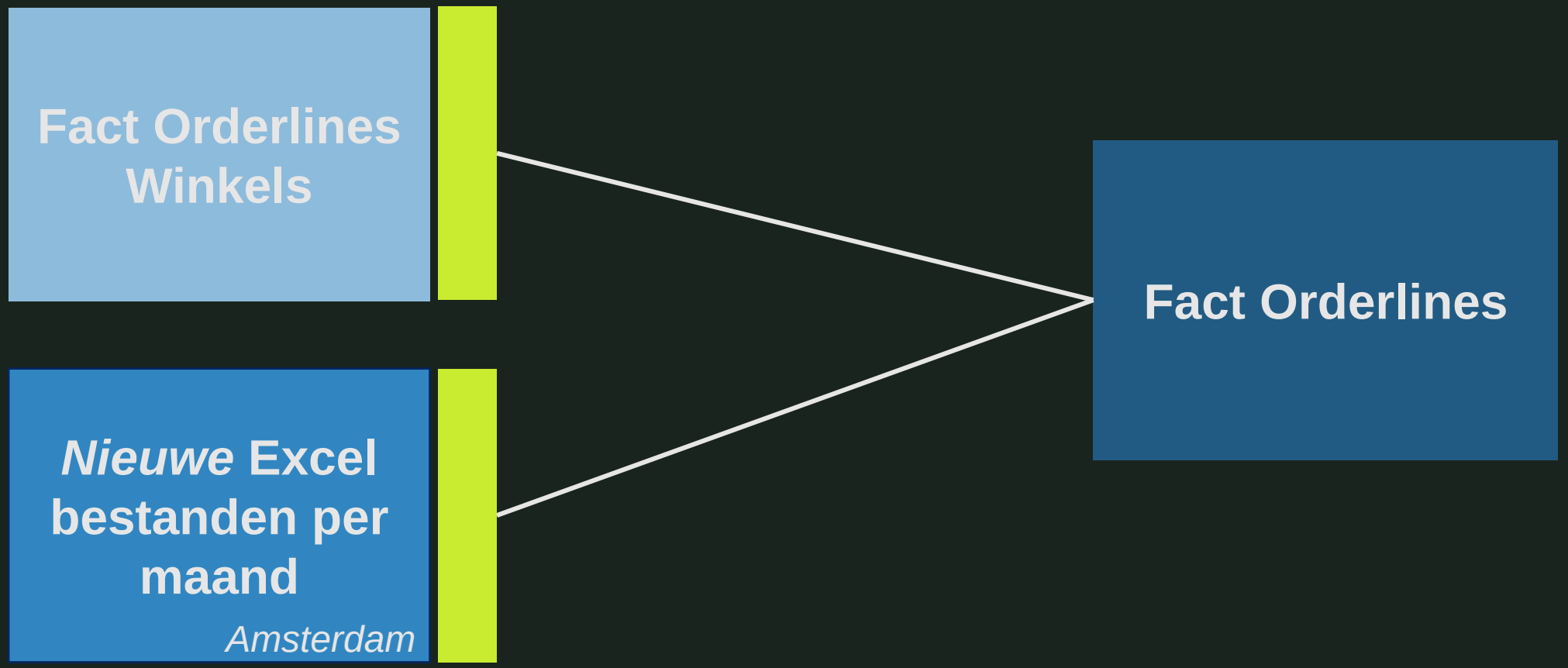
“Amsterdam stuurt elke maand een nieuwe Excel door. Deze data willen wij bewaren en er moet per maand gerapporteerd worden wat er verkocht wordt. Dit is een belangrijke winkel omdat veel toeristen naar deze winkel komen. Hoe kunnen wij zorgen dat wij deze historie opslaan?”

valcon


Naam
Amsterdam 2024 01 30.xlsx
Amsterdam 2024 02 28.xlsx
Amsterdam 2024 03 30.xlsx

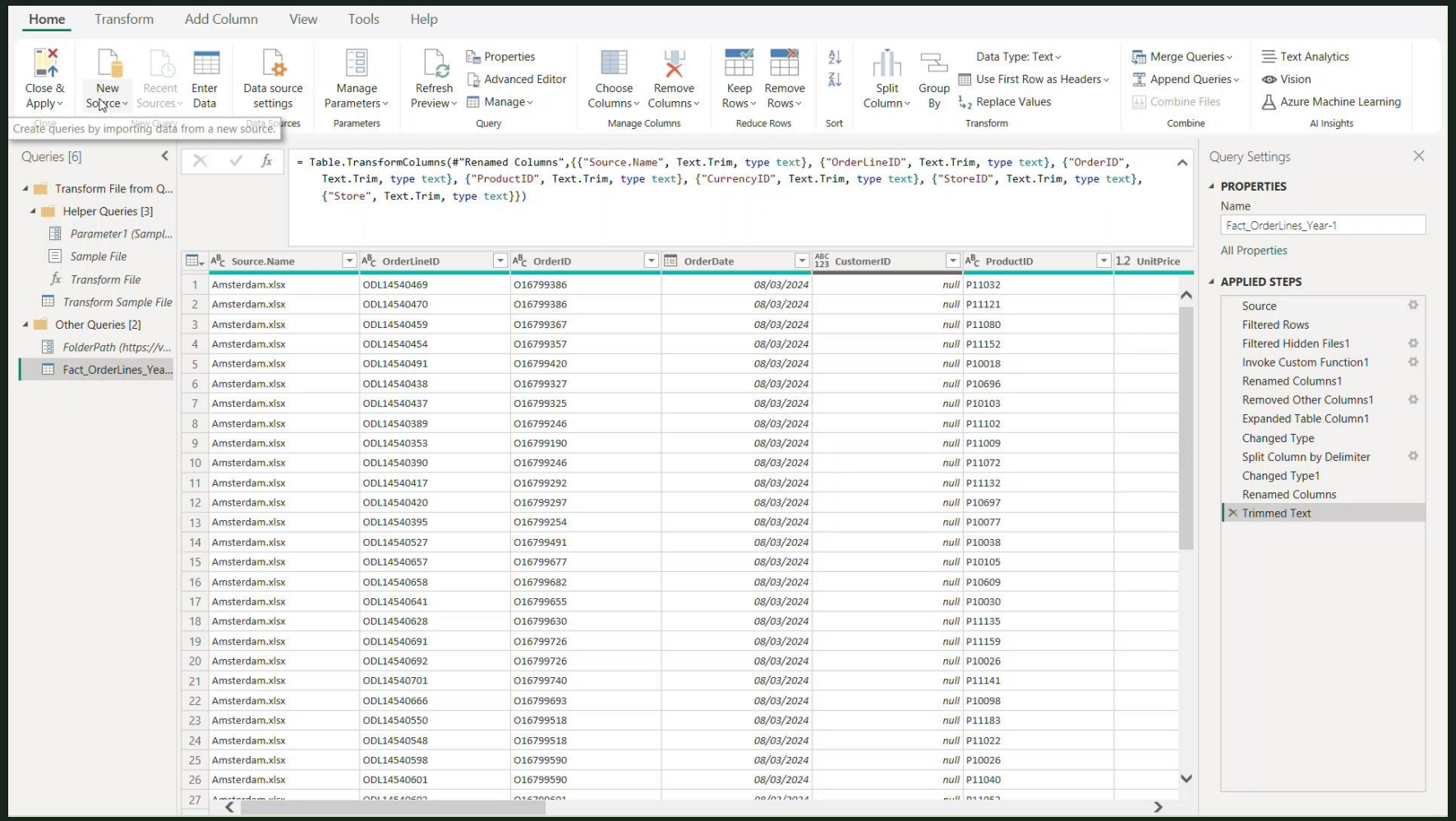
valcon

Best practise



valcon

Nieuwe query + Source date aanmaken



valcon

Status update

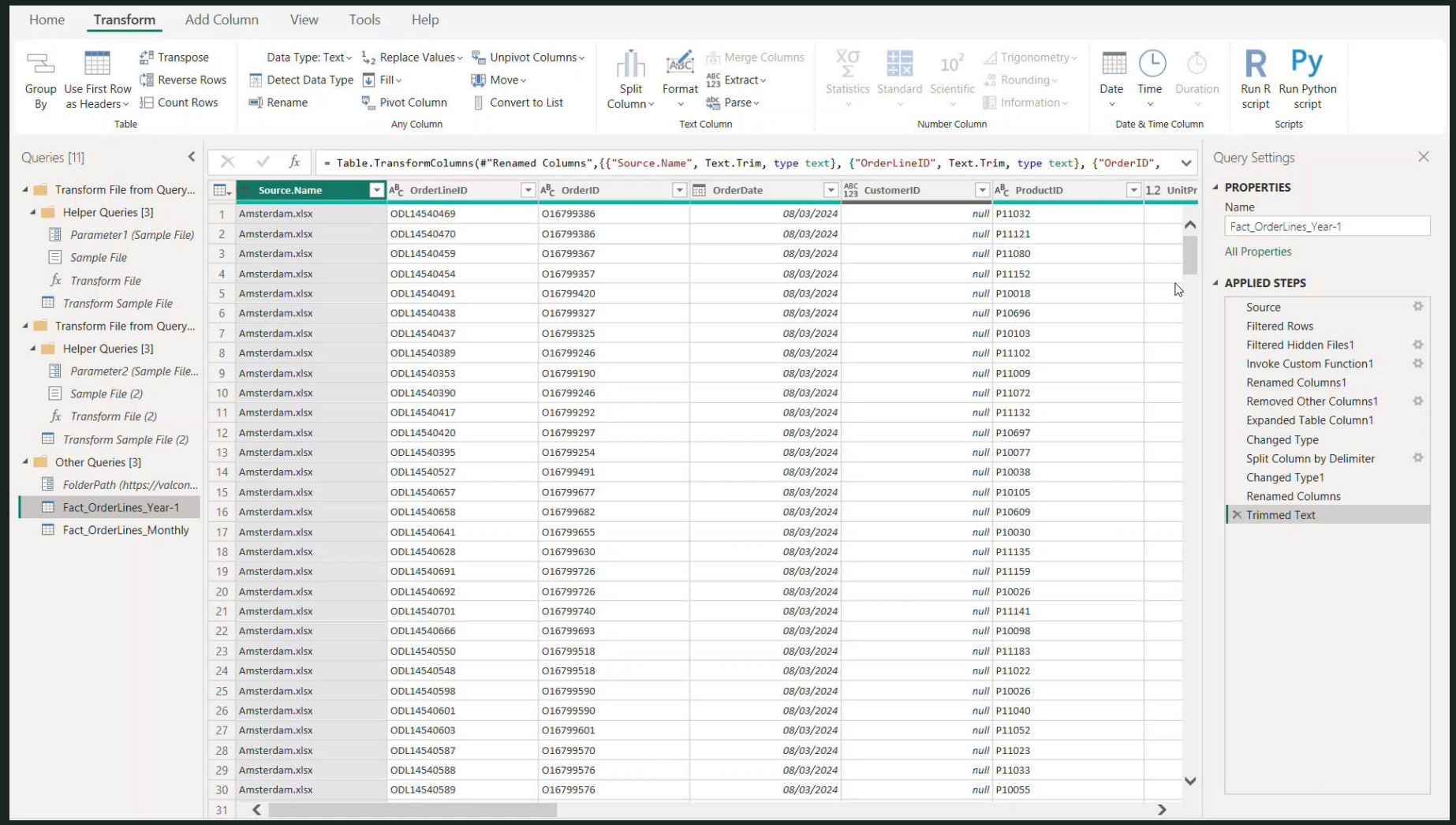
Fact Orderlines Winkels

***Nieuwe* Excel bestanden per maand**

Amsterdam

valcon

Query 1: Source date aanmaken



valcon

Status update

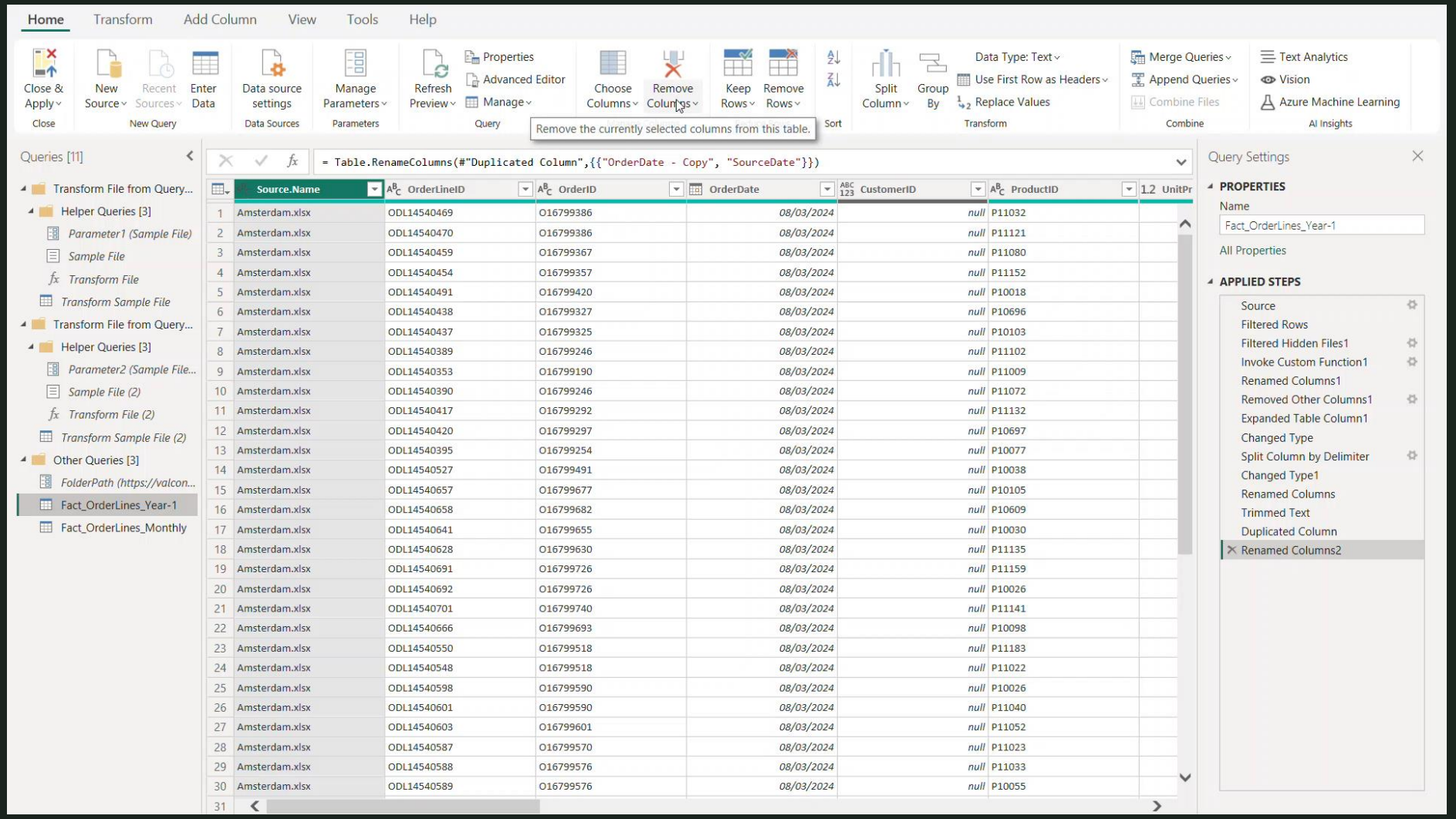
Fact Orderlines Winkels

Nieuwe Excel bestanden per maand

Amsterdam

valcon

Legacy met append



valcon


Data Minimaliseren
valcon

Het management van JungleBoots Factory wil alleen de *nieuwe* cijfers zien en wil dus alleen de laatste 6 maanden in de rapportage.

OrderLineID	OrderID	OrderDate	CustomerID	ProductID	UnitPrice	CurrencyID	Quantity	DiscountID	DiscountAmount	NetPrice	Timestamp	EmployeeID	PaymentID	StoreID
ODL14391498	O16570277	Wednesday, 3 January 2024		P11022	49,95	EUR	1		0	49,95	13:56:07	0	P1	ST01
ODL14391242	O16569895	Wednesday, 3 January 2024		P11022	49,95	EUR	1		0	49,95	14:48:54	0	P1	ST01
ODL14396364	O16577802	Thursday, 4 January 2024		P11022	49,95	EUR	1		0	49,95	11:44:18	0	P1	ST01
ODL14396147	O16577473	Thursday, 4 January 2024		P11022	49,95	EUR	1		0	49,95	09:10:20	0	P1	ST01
ODL14399249	O16582224	Friday, 5 January 2024		P11022	49,95	EUR	1		0	49,95	14:55:01	0	P1	ST01
ODL14401674	O16585922	Saturday, 6 January 2024		P11022	49,95	EUR	1		0	49,95	15:56:35	0	P1	ST01
ODL14403106	O16588126	Saturday, 6 January 2024		P11022	49,95	EUR	1		0	49,95	11:04:28	0	P1	ST01
ODL14403273	O16588374	Sunday, 7 January 2024		P11022	49,95	EUR	1		0	49,95	16:43:08	0	P1	ST01
ODL14407405	O16594753	Monday, 8 January 2024		P11022	49,95	EUR	1		0	49,95	14:10:19	0	P1	ST01
ODL14413955	O16604807	Thursday, 11 January 2024		P11022	49,95	EUR	1		0	49,95	15:04:24	0	P1	ST01
ODL14412083	O16601908	Thursday, 11 January 2024		P11022	49,95	EUR	1		0	49,95	15:38:49	0	P1	ST01
ODL14414466	O16605625	Friday, 12 January 2024		P11022	49,95	EUR	1		0	49,95	12:33:33	0	P1	ST01
ODL14417577	O16610357	Saturday, 13 January 2024		P11022	49,95	EUR	1		0	49,95	13:29:04	0	P1	ST01
ODL14419692	O16613621	Sunday, 14 January 2024		P11022	49,95	EUR	1		0	49,95	14:59:55	0	P1	ST01
ODL14419359	O16613119	Sunday, 14 January 2024		P11022	49,95	EUR	1		0	49,95	13:56:58	0	P1	ST01
ODL14419480	O16613282	Sunday, 14 January 2024		P11022	49,95	EUR	1		0	49,95	13:14:19	0	P1	ST01
ODL14421664	O16616621	Monday, 15 January 2024		P11022	49,95	EUR	1		0	49,95	11:45:00	0	P1	ST01
ODL14427403	O16625552	Wednesday, 17 January 2024		P11022	49,95	EUR	1		0	49,95	13:12:45	0	P1	ST01
ODL14425877	O16623140	Wednesday, 17 January 2024		P11022	49,95	EUR	1		0	49,95	16:24:29	0	P1	ST01
ODL14427140	O16625124	Wednesday, 17 January 2024		P11022	49,95	EUR	1		0	49,95	13:27:53	0	P1	ST01
ODL14426685	O16624428	Wednesday, 17 January 2024		P11022	49,95	EUR	1		0	49,95	10:16:47	0	P1	ST01
ODL14429731	O16629124	Thursday, 18 January 2024		P11022	49,95	EUR	1		0	49,95	09:02:21	0	P1	ST01
ODL14429408	O16628643	Thursday, 18 January 2024		P11022	49,95	EUR	1		0	49,95	12:56:10	0	P1	ST01
ODL14432106	O16632749	Friday, 19 January 2024		P11022	49,95	EUR	1		0	49,95	10:48:44	0	P1	ST01
ODL14432017	O16632611	Friday, 19 January 2024		P11022	49,95	EUR	1		0	49,95	09:55:47	0	P1	ST01
ODL14434678	O16636723	Saturday, 20 January 2024		P11022	49,95	EUR	1		0	49,95	10:00:27	0	P1	ST01
ODL14435521	O16638032	Sunday, 21 January 2024		P11022	49,95	EUR	1		0	49,95	15:40:17	0	P1	ST01
ODL14436141	O16638945	Sunday, 21 January 2024		P11022	49,95	EUR	1		0	49,95	15:54:47	0	P1	ST01
ODL14441110	O16646691	Tuesday, 23 January 2024		P11022	49,95	EUR	1		0	49,95	13:23:48	0	P1	ST01
ODL14441528	O16647364	Tuesday, 23 January 2024		P11022	49,95	EUR	1		0	49,95	14:58:12	0	P1	ST01
ODL14441572	O16647428	Tuesday, 23 January 2024		P11022	49,95	EUR	1		0	49,95	12:01:19	0	P1	ST01
ODL14443452	O16650303	Wednesday, 24 January 2024		P11022	49,95	EUR	1		0	49,95	13:33:11	0	P1	ST01

Gebruik ***variabele*** en een ***M functie*** voor het filteren van de dataset.

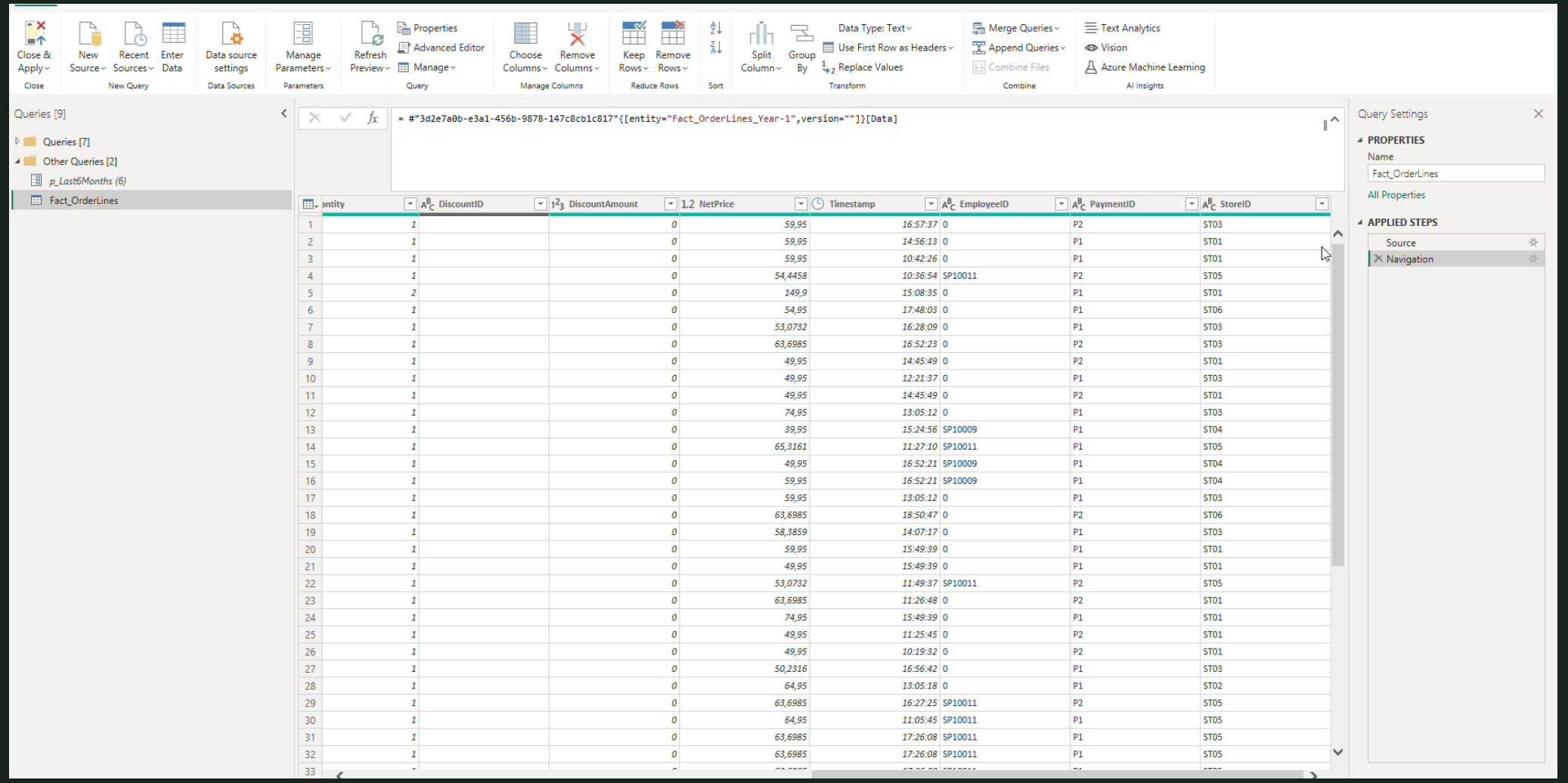
- **Variabele: Last6Months**
- **M functie: List.Max()**

```
= Table.SelectRows(#"Fact_OrderLines_Year-1_",
let last6months = List.Max(#"Fact_OrderLines_Year-1_"[OrderDate])
in
each ([OrderDate] <= Date.AddMonths(Date.From(DateTime.LocalNow()), - p_Last6Months) ))
```

= Table.SelectRows(#"Fact_OrderLines_Year-1_",
let last6months = List.Max(#"Fact_OrderLines_Year-1_"[OrderDate])
in
each ([OrderDate] <= Date.AddMonths(Date.From(DateTime.LocalNow()), - p_Last6Months)))

valcon

Dynamische Filter: Variabele & List.Max Functie



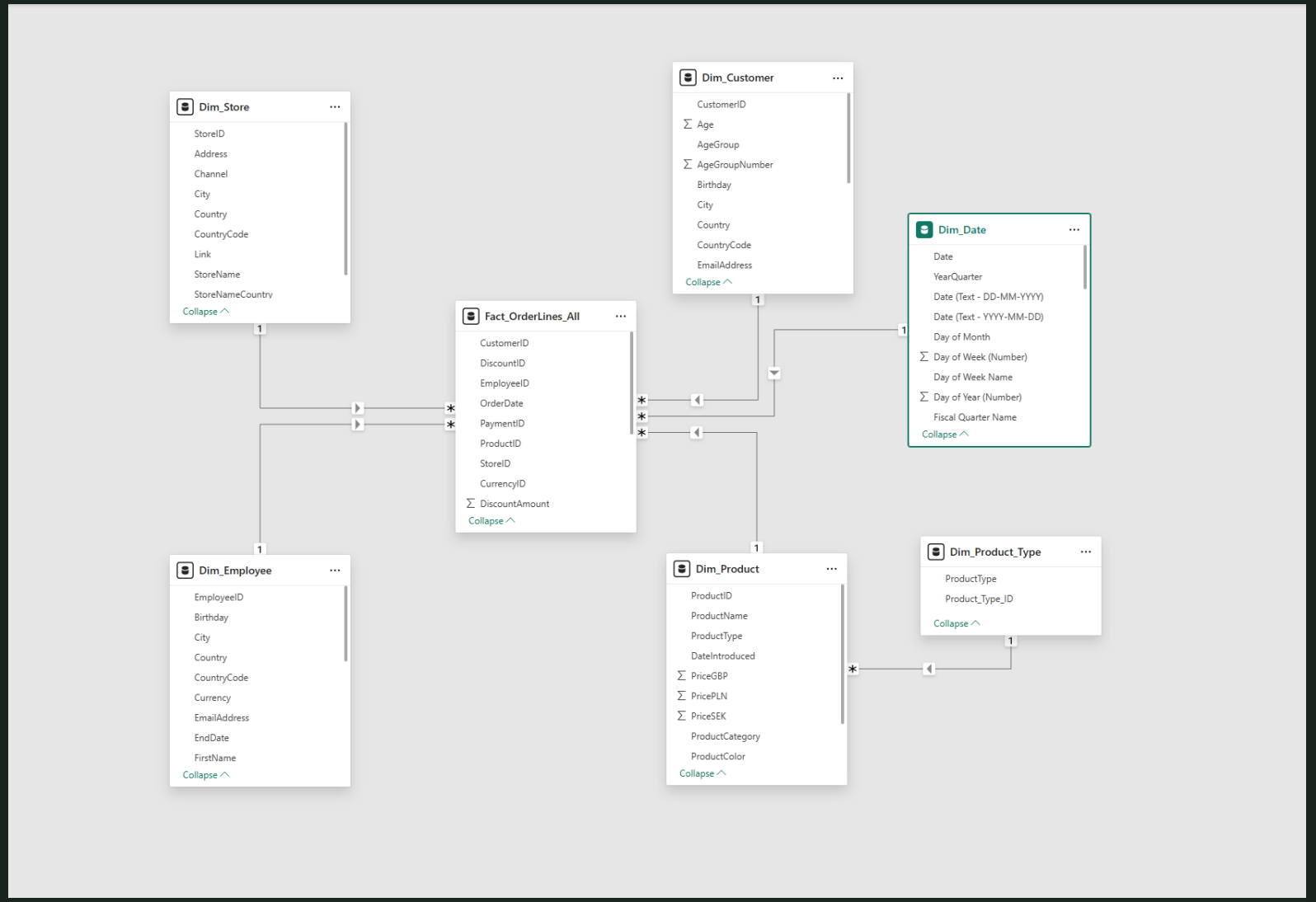
	antity	DiscountID	DiscountAmount	NetPrice	Timestamp	EmployeeID	PaymentID	StoreID
1	1		0	59,95	16:57:37	0	P2	ST03
2	1		0	59,95	14:56:13	0	P1	ST01
3	1		0	59,95	10:42:26	0	P1	ST01
4	1		0	54,4458	10:36:54	SP10011	P2	ST05
5	2		0	149,9	15:08:35	0	P1	ST01
6	1		0	54,95	17:48:03	0	P1	ST06
7	1		0	53,0732	16:28:09	0	P1	ST03
8	1		0	63,6985	16:52:23	0	P2	ST03
9	1		0	49,95	14:45:49	0	P2	ST01
10	1		0	49,95	12:21:37	0	P1	ST03
11	1		0	49,95	14:45:49	0	P2	ST01
12	1		0	74,95	13:05:12	0	P1	ST03
13	1		0	39,95	15:24:56	SP10009	P1	ST04
14	1		0	65,3161	11:27:10	SP10011	P1	ST05
15	1		0	49,95	16:52:21	SP10009	P1	ST04
16	1		0	59,95	16:52:21	SP10009	P1	ST04
17	1		0	59,95	13:05:12	0	P1	ST03
18	1		0	63,6985	18:50:47	0	P2	ST06
19	1		0	58,3859	14:07:17	0	P1	ST03
20	1		0	59,95	15:49:39	0	P1	ST01
21	1		0	49,95	15:49:39	0	P1	ST01
22	1		0	53,0732	11:49:37	SP10011	P2	ST05
23	1		0	63,6985	11:26:48	0	P2	ST01
24	1		0	74,95	15:49:39	0	P1	ST01
25	1		0	49,95	11:25:45	0	P2	ST01
26	1		0	49,95	10:19:32	0	P2	ST01
27	1		0	50,2316	16:56:42	0	P1	ST03
28	1		0	64,95	13:05:18	0	P1	ST02
29	1		0	63,6985	16:27:25	SP10011	P2	ST05
30	1		0	64,95	11:05:45	SP10011	P1	ST05
31	1		0	63,6985	17:26:08	SP10011	P1	ST05
32	1		0	63,6985	17:26:08	SP10011	P1	ST05

valcon


valcon


valcon

Semantisch model



valcon

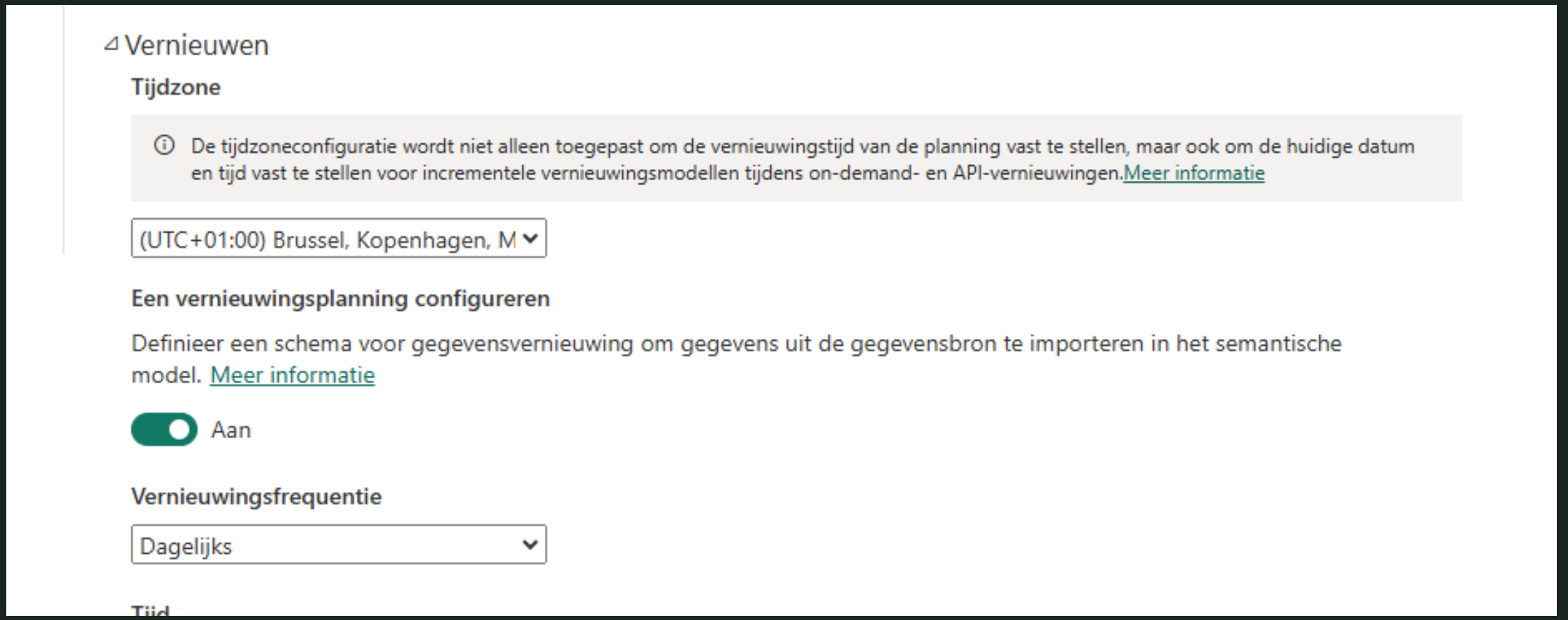
Data verversen

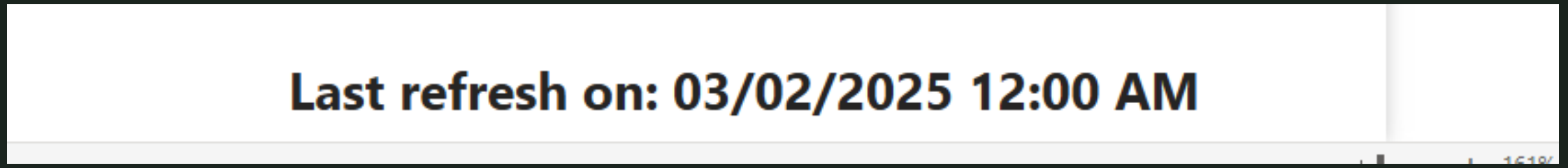
valcon

Verversen

- Frequentie van oplevering
- Data dagelijks verversen instellen
- Datum verversing en Source date tonen op Rapport

valcon


Vernieuwen
Tijdzone
De tijdzoneconfiguratie wordt niet alleen toegepast om de vernieuwingstijd van de planning vast te stellen, maar ook om de huidige datum en tijd vast te stellen voor incrementele vernieuwingsmodellen tijdens on-demand- en API-vernieuwingen.Meer informatie
(UTC+01:00) Brussel, Kopenhagen, M
Een vernieuwingsplanning configureren
Definieer een schema voor gegevensvernieuwing om gegevens uit de gegevensbron te importeren in het semantische model. Meer informatie
Aan
Vernieuwingsfrequentie
Dagelijks
Tijd


Last refresh on: 03/02/2025 12:00 AM

valcon

Visualisatie

valcon

JungleBoots Factory

Last Source Date: 2/2/2025

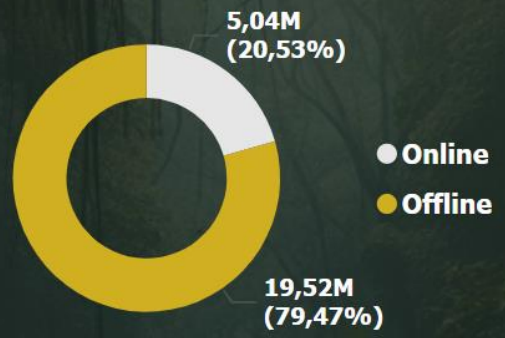
Sales Overview

Sales

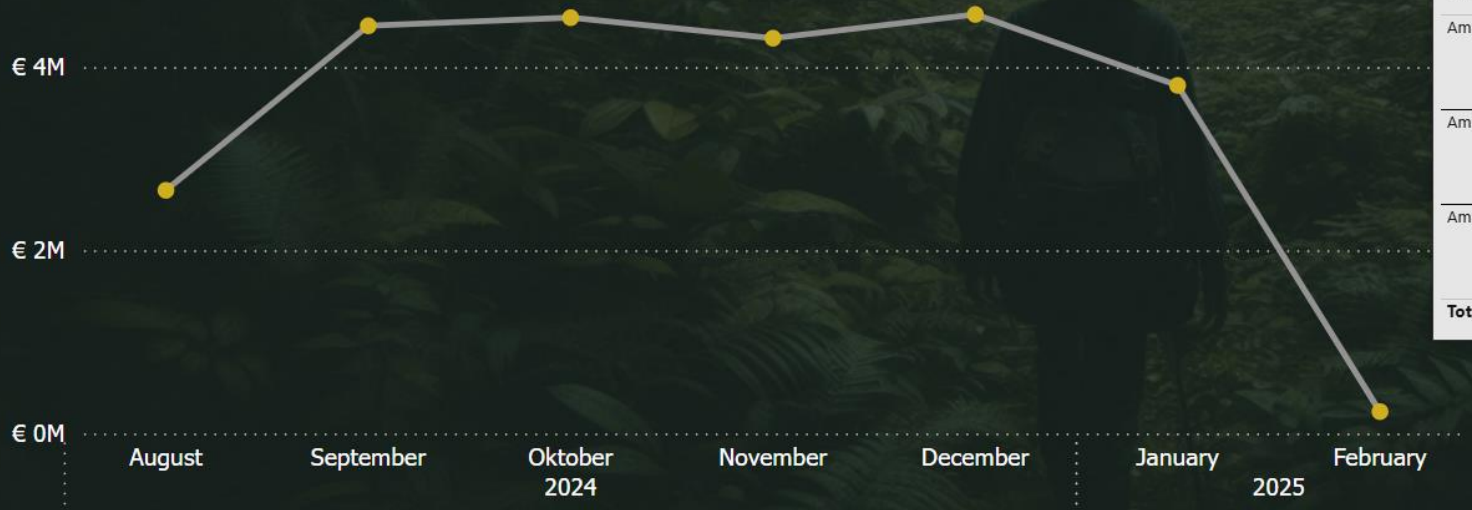
€ 24,57M

Growth -17,19%

Sales Online and Offline

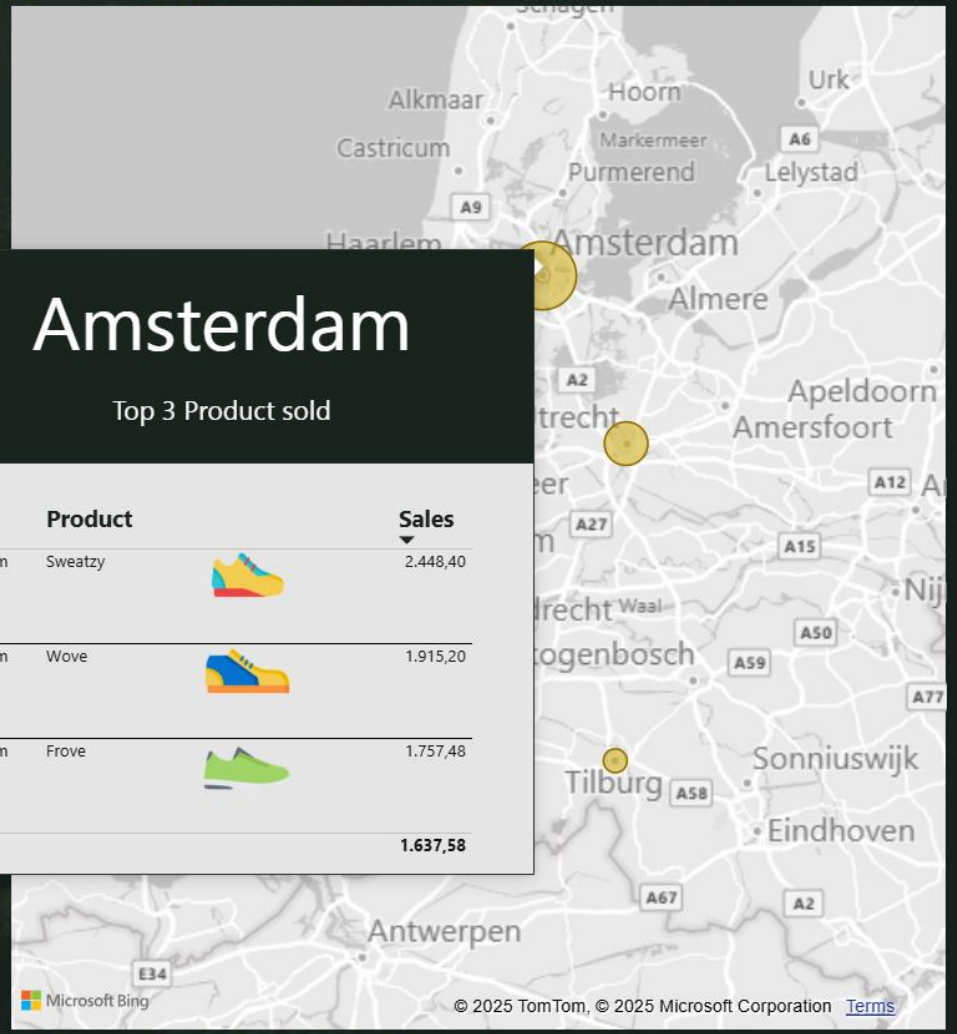


Sales Yearly



Stores

Amsterdam Utrecht Den Haag Tilburg



Amsterdam

Top 3 Product sold

Store	Product	Sales
Amsterdam	Sweatzy	2.448,40
Amsterdam	Wove	1.915,20
Amsterdam	Frove	1.757,48
Totaal		**1.637,58**

Last Refresh: 06/02/2025 14:00:49 +01:00

Samenvatting van handvatten, tips en trucs

valcon

1. Begin met alles te verzamelen naar een (Sharepoint) map
   - Let op naamgeving van de bestanden
2. Laad de gehele map(pen) in via een parameter
   - Inclusief historiseren van data
3. Begin met Splitsen van kolommen
4. Verwijder onzichtbare spaties → Text.Trim
5. Voeg queries samen → Append
6. Maak een semantisch model inclusief datum tabel
7. Visualiseer waar *echt* om is gevraagd (incl. continue feedbackloop)

- **Maak goede afspraken met je data owners!**
- **Investeer in Power BI en Power Query kennis**

valcon

Power BI met Excel als bron | Stappenplan

Afspraken

- [ ] SLA met Data Owners
- [ ] Frequentie Bepalen
- [ ] Oplaglocatie vaststellen

Verzamelen

- [ ] Aspraken maken
- [ ] Maak een shared opslag voor Excel bestanden
- [ ] Bepaal de toegang voor data owners.
- [ ] Naamgeving gelijk houden en controle van de bestanden.
- [ ] Structuur afspreken

Inladen

- [ ] Gebruik Parameters
- [ ] Kolomstructuur aanhouden
- [ ] Data consolideren

Data Cleansing

- [ ] Kolommen Splisten met Split functie
- [ ] Verwijder onzichtbare spaties. → Text.Trim
- [ ] Samenvoegen. → Append

Model

- [ ] Tabellen combineren voor consolidatie.
- [ ] Gebruik source datum kolommen.
- [ ] Data en kolommen minimaliseren
- [ ] Van Plat naar Ster model

Visualiseren

- [ ] Verversdatum op rapport
- [ ] Source Datum op rapport
- [ ] Dagelijks verversen instellen tbv beide datums
- [ ] Creëren van benodigde visualisaties

with valcon

Geanpaul Mercera

Noa van Hunen

Dankjewel voor jullie aandacht

valcon