# Power BI - Data Modelling - p1

20 October 2024 12:35

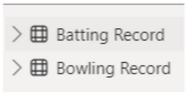
## Introduction To Data Modelling - p1

### dd/mm/yyyy ---> 1st January 2020

| OrderDate | StockDate |
|---|---|
| 01 January 2020 | 29 October 2019 |
| 02 January 2020 | 18 December 2019 |
| 02 January 2020 | 09 October 2019 |
| 03 January 2020 | 03 October 2019 |
| 03 January 2020 | 24 October 2019 |

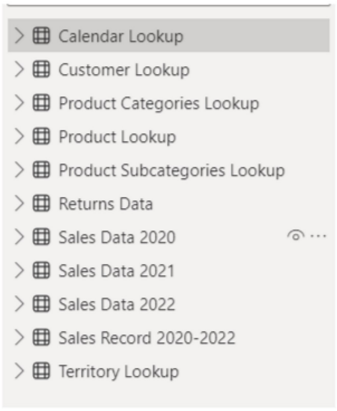
We have to fix it

### Step 1 : Delete all the pre-built relations added automatically.




Delete [Web Scrapping understood , now no need]

### This should be my final cleaned dataset.

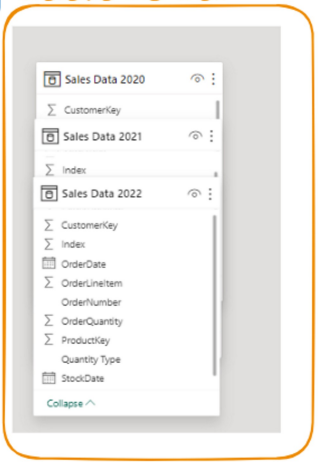


Put the data Table in this order [having Dimension Table Above Fact Table].


Dimension Table

Fact Table

Not useful but can't delete as well




| ProductName | Sum of OrderQuantity | Sum of ReturnQuantity |
|---|---|---|
| All-Purpose Bike Stand | 84174 | 1828 |
| AWC Logo Cap | 84174 | 1828 |
| Bike Wash - Dissolver | 84174 | 1828 |
| Cable Lock | 84174 | 1828 |
| Chain | 84174 | 1828 |
| Classic Vest, L | 84174 | 1828 |
| Classic Vest, M | 84174 | 1828 |
| Classic Vest, S | 84174 | 1828 |
| Fender Set - Mountain | 84174 | 1828 |
| Front Brakes | 84174 | 1828 |
| Front Derailleur | 84174 | 1828 |
| Full-Finger Gloves, L | 84174 | 1828 |
| Full-Finger Gloves, M | 84174 | 1828 |
| Full-Finger Gloves, S | 84174 | 1828 |
| Half-Finger Gloves, L | 84174 | 1828 |
| Half-Finger Gloves, M | 84174 | 1828 |
| Half-Finger Gloves, S | 84174 | 1828 |
| Headlights - Dual-Beam | 84174 | 1828 |
| **Total** | **84174** | **1828** |

single Filter

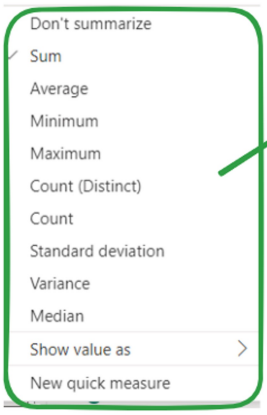

Filter Flow : Dimension --- Fact

1 : Many Relationship [Cardinality]

Dimension : Where you will categorical data which is uniquely present.

Fact Table : Which stores Transaction record [Statistics - apply] [Numerical Data]

Total Price
Total Order Qty.
Profit
Return Rate
Weekend Order




Aggregation [Stats]

## Create relationship

Select tables and columns that are related.

Product Lookup

| ProductKey | ProductSubcategoryKey | ProductSKU | SKU Type | ProductName | ModelName | |
|---|---|---|---|---|---|---|
| 310 | 2 | BK-R93R-62 | BK | Road-150 Red, 62 | Road-150 | This bike is rid |
| 311 | 2 | BK-R93R-44 | BK | Road-150 Red, 44 | Road-150 | This bike is rid |
| 312 | 2 | BK-R93R-48 | BK | Road-150 Red, 48 | Road-150 | This bike is rid |

Returns Data

| ReturnDate | TerritoryKey | ProductKey | ReturnQuantity |
|---|---|---|---|
| 18 January 2020 | 9 | 312 | 1 |
| 19 February 2020 | 9 | 311 | 1 |
| 13 March 2020 | 9 | 350 | 1 |

Cardinality: One to many (1:*)

Cross filter direction: Single

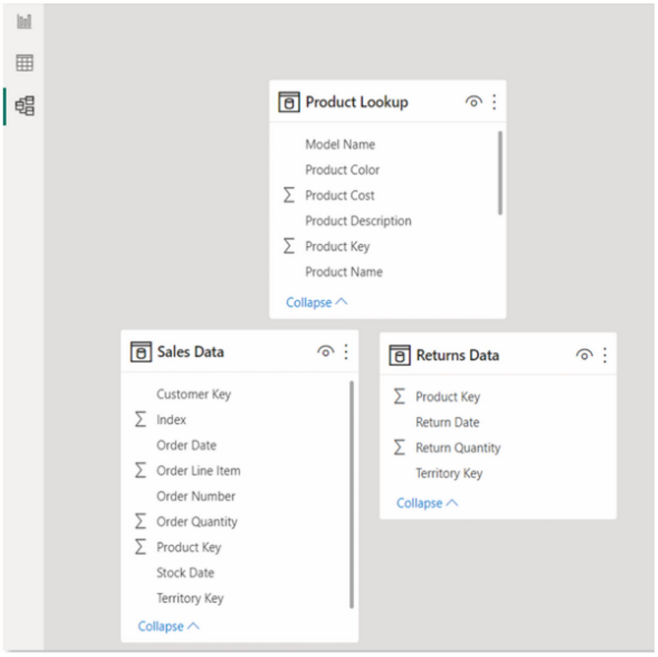
Make this relationship active

Assume referential integrity

Apply security filter in both directions

OK Cancel

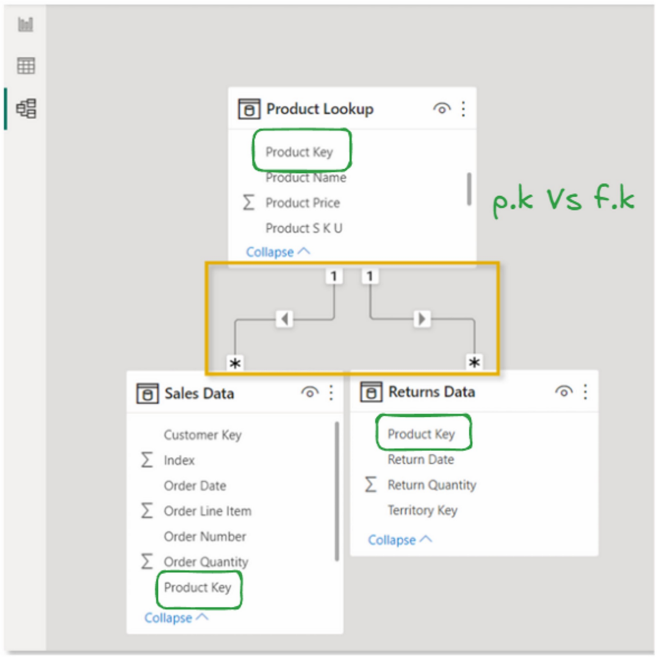
# WHAT IS A DATA MODEL?




This IS NOT a data model

- This is a collection of independent tables, which share no connections or relationships
- If you tried to visualize Orders and Returns by Product, this is what you'd get

| ProductName | OrderQuantity | ReturnQuantity |
|---|---|---|
| All-Purpose Bike Stand | 84,174 | 1,828 |
| AWC Logo Cap | 84,174 | 1,828 |
| Bike Wash - Dissolver | 84,174 | 1,828 |
| Cable Lock | 84,174 | 1,828 |
| Chain | 84,174 | 1,828 |
| Classic Vest, L | 84,174 | 1,828 |
| Classic Vest, M | 84,174 | 1,828 |
| Classic Vest, S | 84,174 | 1,828 |
| Fender Set - Mountain | 84,174 | 1,828 |
| **Total** | **84,174** | **1,828** |




p.k Vs f.k

# This IS a data model! 😊

- The tables are connected via relationships, based on a common field (Product Key)
- Now Sales and Returns data can be filtered using fields from the Product Lookup table!

| ProductName | OrderQuantity | ReturnQuantity |
|---|---|---|
| All-Purpose Bike Stand | 234 | 8 |
| AWC Logo Cap | 4,151 | 46 |
| Bike Wash - Dissolver | 1,706 | 25 |
| Classic Vest, L | 182 | 4 |
| Classic Vest, M | 182 | 7 |
| Classic Vest, S | 157 | 8 |
| Fender Set - Mountain | 3,960 | 54 |
| Half-Finger Gloves, L | 840 | 18 |
| Half-Finger Gloves, M | 918 | 16 |
| **Total** | **84,174** | **1,828** |

# DATABASE NORMALIZATION

Normalization is the process of organizing the tables and columns in a relational database to reduce redundancy and preserve data integrity. It's commonly used to:

- Eliminate redundant data to decrease table sizes and improve processing speed & efficiency.
- Minimize errors and anomalies from data modifications (inserting, updating or deleting records).
- Simplify queries and structure the database for meaningful analysis.

In a normalized database, each table should serve a distinct and specific purpose (i.e. product information, transaction records, customer attributes, store details, etc.)

| date | product_id | quantity | product_brand | product_name | product_sku | product_weight |
|---|---|---|---|---|---|---|
| 1/1/1997 | 869 | 5 | Nationeel | Nationeel Grape Fruit Roll | 52382137179 | 17 |
| 1/7/1997 | 869 | 2 | Nationeel | Nationeel Grape Fruit Roll | 52382137179 | 17 |
| 1/3/1997 | 1 | 4 | Washington | Washington Berry Juice | 90748583674 | 8.39 |
| 1/1/1997 | 1472 | 3 | Fort West | Fort West Fudge Cookies | 37276054024 | 8.28 |
| 1/6/1997 | 1472 | 2 | Fort West | Fort West Fudge Cookies | 37276054024 | 8.28 |
| 1/5/1997 | 2 | 4 | Washington | Washington Mango Drink | 96516502499 | 7.42 |
| 1/1/1997 | 76 | 4 | Red Spade | Red Spade Sliced Chicken | 62054644227 | 18.1 |
| 1/1/1997 | 76 | 2 | Red Spade | Red Spade Sliced Chicken | 62054644227 | 18.1 |
| 1/5/1997 | 3 | 2 | Washington | Washington Strawberry Drink | 58427771925 | 13.1 |
| 1/7/1997 | 3 | 2 | Washington | Washington Strawberry Drink | 58427771925 | 13.1 |
| 1/1/1997 | 320 | 3 | Excellent | Excellent Cranberry Juice | 36570182442 | 16.4 |

→ Models that aren't normalized contain redundant, duplicate data. In this case, all of the product-specific fields could be stored in a separate table containing a unique record for each product id

→ This may not seem critical now, but minor inefficiencies can become major problems at scale!

# FACT & DIMENSION TABLES

Data models generally contain two types of tables: fact ("data") tables, and dimension ("lookup") tables:

- Fact tables contain numerical values or metrics used for summarization (sales, orders, transactions, pageviews, etc.)

- Dimension tables contain descriptive attributes used for filtering or grouping (products, customers, dates, stores, etc.)

| date | product_id | quantity |
|---|---|---|
| 1/1/1997 | 869 | 5 |
| 1/1/1997 | 1472 | 3 |
| 1/1/1997 | 76 | 4 |
| 1/1/1997 | 320 | 3 |
| 1/1/1997 | 4 | 4 |
| 1/1/1997 | 952 | 4 |
| 1/1/1997 | 1222 | 4 |
| 1/1/1997 | 517 | 4 |
| 1/1/1997 | 1359 | 4 |
| 1/1/1997 | 357 | 4 |
| 1/1/1997 | 1426 | 5 |
| 1/1/1997 | 190 | 4 |
| 1/1/1997 | 367 | 4 |
| 1/1/1997 | 250 | 5 |
| 1/1/1997 | 600 | 4 |
| 1/1/1997 | 702 | 5 |

This Fact table contains quantity values, along with date and product_id fields

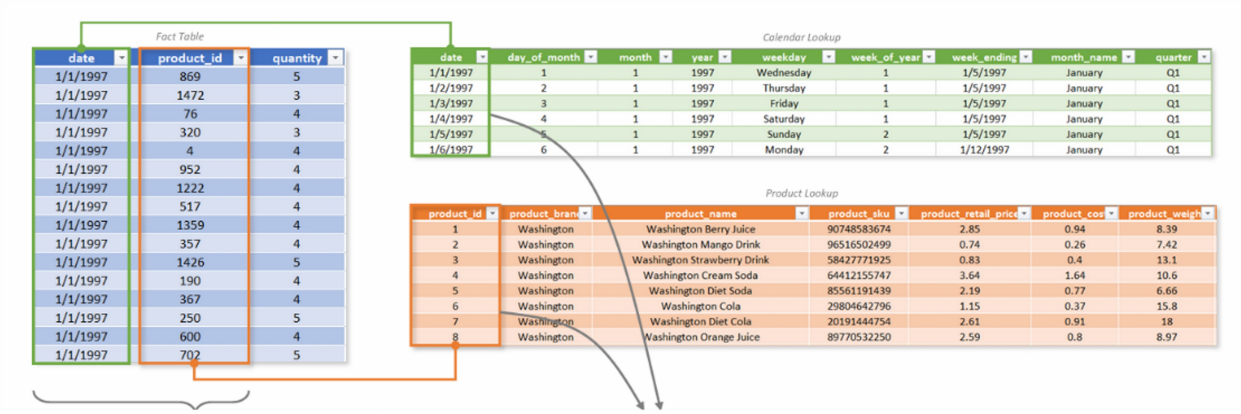
| date | day_of_month | month | year | weekday | week_of_year | week_ending | month_name | quarter |
|---|---|---|---|---|---|---|---|---|
| 1/1/1997 | 1 | 1 | 1997 | Wednesday | 1 | 1/5/1997 | January | Q1 |
| 1/2/1997 | 2 | 1 | 1997 | Thursday | 1 | 1/5/1997 | January | Q1 |
| 1/3/1997 | 3 | 1 | 1997 | Friday | 1 | 1/5/1997 | January | Q1 |
| 1/4/1997 | 4 | 1 | 1997 | Saturday | 1 | 1/5/1997 | January | Q1 |
| 1/5/1997 | 5 | 1 | 1997 | Sunday | 2 | 1/5/1997 | January | Q1 |
| 1/6/1997 | 6 | 1 | 1997 | Monday | 2 | 1/12/1997 | January | Q1 |

This Calendar Lookup table contains attributes about each date (month, year, quarter, etc.)

| product_id | product_brand | product_name | product_sku | product_retail_price | product_cost | product_weight |
|---|---|---|---|---|---|---|
| 1 | Washington | Washington Berry Juice | 90748583674 | 2.85 | 0.94 | 8.39 |
| 2 | Washington | Washington Mango Drink | 96516502499 | 0.74 | 0.26 | 7.42 |
| 3 | Washington | Washington Strawberry Drink | 58427771925 | 0.83 | 0.4 | 13.1 |
| 4 | Washington | Washington Cream Soda | 64412155747 | 3.64 | 1.64 | 10.6 |
| 5 | Washington | Washington Diet Soda | 85561191439 | 2.19 | 0.77 | 6.66 |
| 6 | Washington | Washington Cola | 29804642796 | 1.15 | 0.37 | 15.8 |
| 7 | Washington | Washington Diet Cola | 20191444754 | 2.61 | 0.91 | 18 |
| 8 | Washington | Washington Orange Juice | 89770532250 | 2.59 | 0.8 | 8.97 |

This Product Lookup table contains attributes about each product_id (brand, SKU, price, etc.)

# PRIMARY & FOREIGN KEYS



*Fact Table*

| date | product_id | quantity |
|---|---|---|
| 1/1/1997 | 869 | 5 |
| 1/1/1997 | 1472 | 3 |
| 1/1/1997 | 76 | 4 |
| 1/1/1997 | 320 | 3 |
| 1/1/1997 | 4 | 4 |
| 1/1/1997 | 952 | 4 |
| 1/1/1997 | 1222 | 4 |
| 1/1/1997 | 517 | 4 |
| 1/1/1997 | 1359 | 4 |
| 1/1/1997 | 357 | 4 |
| 1/1/1997 | 1426 | 5 |
| 1/1/1997 | 190 | 4 |
| 1/1/1997 | 367 | 4 |
| 1/1/1997 | 250 | 5 |
| 1/1/1997 | 600 | 4 |
| 1/1/1997 | 702 | 5 |

*Calendar Lookup*

| date | day_of_month | month | year | weekday | week_of_year | week_ending | month_name | quarter |
|---|---|---|---|---|---|---|---|---|
| 1/1/1997 | 1 | 1 | 1997 | Wednesday | 1 | 1/5/1997 | January | Q1 |
| 1/2/1997 | 2 | 1 | 1997 | Thursday | 1 | 1/5/1997 | January | Q1 |
| 1/3/1997 | 3 | 1 | 1997 | Friday | 1 | 1/5/1997 | January | Q1 |
| 1/4/1997 | 4 | 1 | 1997 | Saturday | 1 | 1/5/1997 | January | Q1 |
| 1/5/1997 | 5 | 1 | 1997 | Sunday | 2 | 1/5/1997 | January | Q1 |
| 1/6/1997 | 6 | 1 | 1997 | Monday | 2 | 1/12/1997 | January | Q1 |

*Product Lookup*

| product_id | product_brand | product_name | product_sku | product_retail_price | product_cost | product_weight |
|---|---|---|---|---|---|---|
| 1 | Washington | Washington Berry Juice | 90748583674 | 2.85 | 0.94 | 8.39 |
| 2 | Washington | Washington Mango Drink | 96516502499 | 0.74 | 0.26 | 7.42 |
| 3 | Washington | Washington Strawberry Drink | 58427771925 | 0.83 | 0.4 | 13.1 |
| 4 | Washington | Washington Cream Soda | 64412155747 | 3.64 | 1.64 | 10.6 |
| 5 | Washington | Washington Diet Soda | 85561191439 | 2.19 | 0.77 | 6.66 |
| 6 | Washington | Washington Cola | 29804642796 | 1.15 | 0.37 | 15.8 |
| 7 | Washington | Washington Diet Cola | 20191444754 | 2.61 | 0.91 | 18 |
| 8 | Washington | Washington Orange Juice | 89770532250 | 2.59 | 0.8 | 8.97 |

**These are foreign keys (FK)**

They contain multiple instances of each value, and relate to primary keys in dimension tables

**These are primary keys (PK)**

They uniquely identify each row of the table, and relate to foreign keys in fact tables