- Data Modeling is the process of creating data model.

- Data model defines the data structure, properties, and relation.

- It is used to connect multiple data sources using a relationship.

- Power BI allows relationships to be built from tables with different data sources

- A relationship defines how data sources are connected with each other and you can create interesting data visualizations on multiple data sources.

- Generally, Smaller data model is better because it will perform faster and will be simpler to use.

- Smaller data model is comprised of fewer tables and fewer columns in each table that the user can see
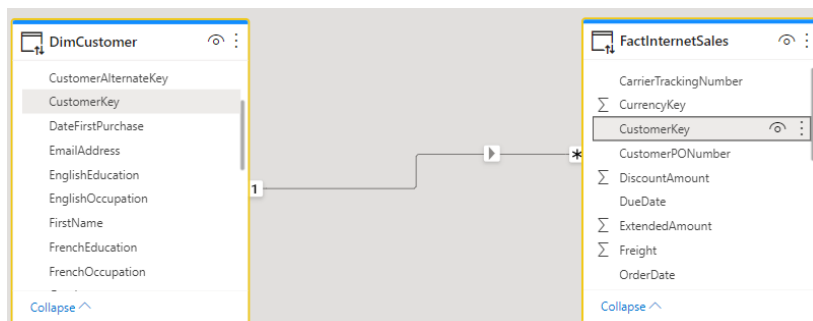
Example:
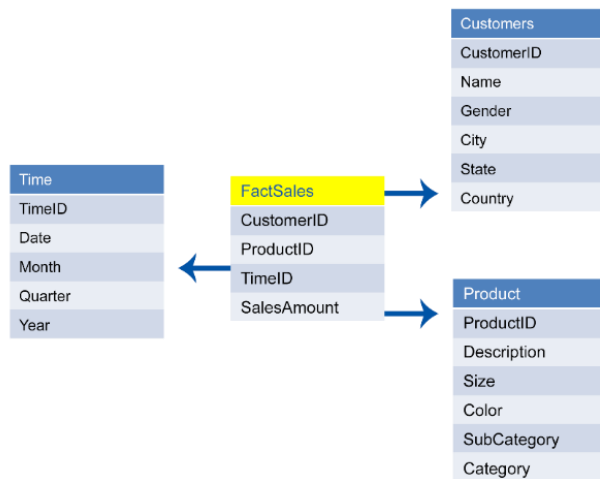
Joins and Relationship

Primary Key: A field (or fields) used to identify a row Uniquely.

Foreign Key: A field (or fields) on the many side of a one-to-many relationship between tables that relates to the primary key of the other table.

Department

ID	Name	Location
101	Sales	Block A
102	IT	Block B

Employee

ID	Name	Salary
1001	John	10000
1002	Allen	15000
1003	Smith	18000

ID	Name	Location
101	Sales	Block A
102	IT	Block B

ID	Name	Salary	Department
1001	John	10000	101
1002	Allen	7000	102
1003	Smith	15000	101

DimCustomer
- CustomerAlternateKey
- CustomerKey
- DateFirstPurchase
- EmailAddress
- EnglishEducation
- EnglishOccupation
- FirstName
- FrenchEducation
- FrenchOccupation

Collapse ∧

FactInternetSales
- CarrierTrackingNumber
- Σ CurrencyKey
- CustomerKey
- CustomerPONumber
- Σ DiscountAmount
- DueDate
- Σ ExtendedAmount
- Σ Freight
- OrderDate

Collapse ∧

Star Schema Design

- Star Schema is the Data modelling approach widely used in relational Datawarehouse
- Star Schema requires data modelers to **classify** their model **tables as either *dimension* or *fact*.**

Fact Table:

Fact tables store observations or events, and can be sales orders, stock balances, exchange rates, temperatures, etc

- It is a table which contains Measures which need to be analyzed.
  Example: Sales Amount Transactions,Units,Cost
- It contains numeric information which can be aggregated for analysis and usually sliceable
- Each fact table relates to one or more dimensions tables.

Example:

FactSales
CustomerId
ProductId
TimeId
SalesAmount

SalesQty

Dimension table:

Dimension tables describe business entities like products, people, places, and concepts including time itself.

- It contains the attributes based on which you need to summarize or analyze the data.
- Example: Product, Customer, Geo
- It contains textual, descriptive information
- It doesn't contain measurable information

Example:

Product
ProductId
ProductName
Size
Color
SubCategory
Category

Note: Generally, dimension tables contain a relatively small number of rows. Fact tables, on the other hand, can contain a very large number of rows and continue to grow over time.

Star Schema:
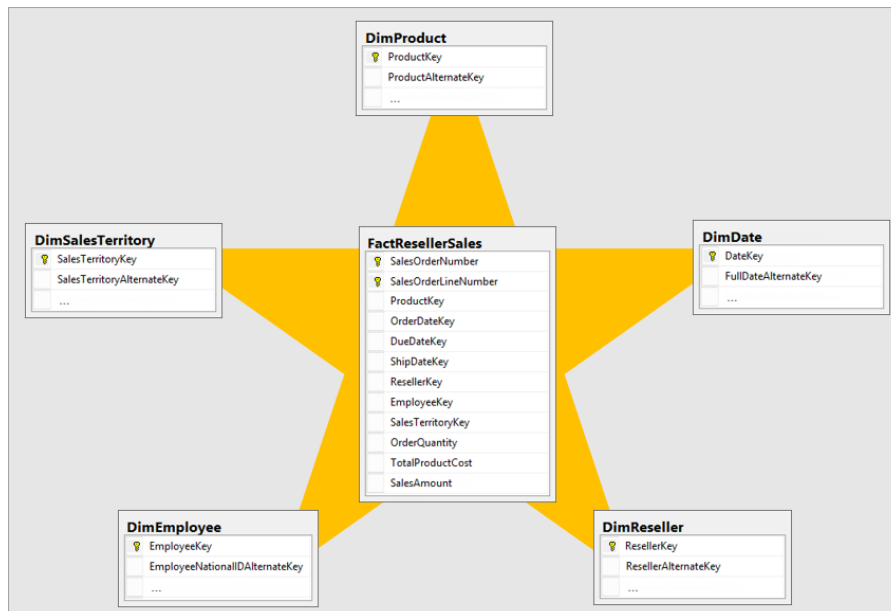
- In PowerBI, Model designed using star schema are optimized for performance and usability.
- Power BI report visual generates a query that is sent to the Power BI model

- These queries are used to filter, group, and summarize model data.
- A well-designed model, then, is one that provides tables for filtering and grouping, and tables for summarizing.
- This design fits well with star schema principles:
  - Dimension tables support *filtering* and *grouping*
  - Fact tables support *summarization*

Steps for Data Modeling:

1. Identify Tables and Relationship Between them.
2. Identify required columns and Remove Unnecessary columns.
3. Define Metadata (Table and Column Properties)

Working with Data Model

- You need to assemble **queries** into a coherent data model and ensure that the underlying structure is clean, ready for use, and contains all the metrics that your dashboards require.
- For data modeling there are several techniques like
  - Specifying data types.
  - Formatting data in the data model.
  - Categorizing data.
  - Adding "sort by" columns that ensure the correct sort order in dashboard elements.
  - Establishing relationships between the tables so that Power BI Desktop understands how the data in one table is linked to the data contained in another table.

**Example:** Create Data Model for Adventureworks Database Tables

Understanding Importance of Relationship:

**Lab 1: Create Model Relationship Between Sales and Product Table**

1. Open file D:\PowerBI-Learn\ClassLabs\Model Data\AdworksSales-starter.pbix
2. Save File as D:\PowerBI-Learn\ClassLabs\Model Data\AdworksSales-final.pbix
3. In Power BI Desktop, at the left, select the **Model** view icon.

4. If you don't see all seven tables, scroll horizontally to the right, and then drag and arrange the tables more closely together so they can all be seen at the same time.

5. To return to Report view, at the left, select the **Report** view icon.
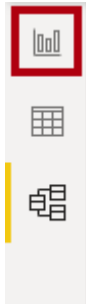
6. To view all table fields, in the **Data** pane, right-click an empty area, and then select **Expand All**.

7. To create a table visual, in the **Data** pane, from inside the **Product** table, check the **Category** field.

8. To add another column to the table, in the **Data** pane, check the **Sales | Sales Amount** field.

9. Observe that aggregates Sales value are not filtered based on Category

10. Go to Model View and observe that the **relationship** between Product and Sales is **not defined**.

11. Sales Table→Right Click→Manage Relationship→Click on New→

12. In the **Create Relationship** window, in the first dropdown list, select the **Product** table.
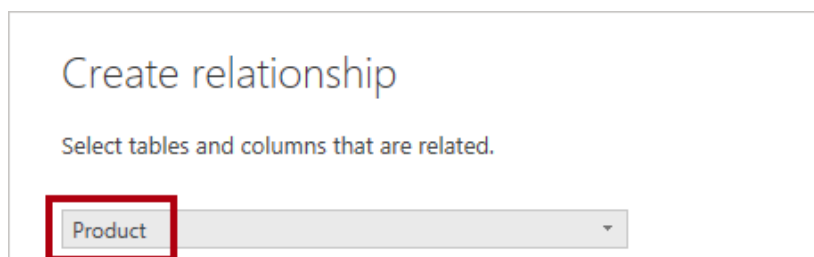
13. In the second dropdown list (beneath the **Product** table grid), select the **Sales** table.

14. Notice the **ProductKey** columns in each table have been automatically selected.

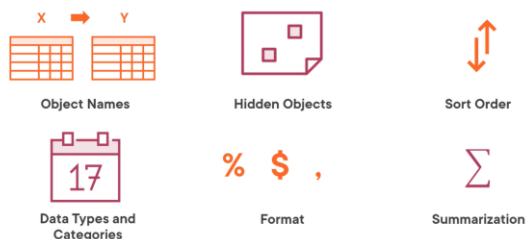15. In the **Cardinality** dropdown list, notice that **One To Many (1:*)** is selected.

Click OK➔Close

- Observe the changes in visual to show Product Category wise Sales

Working With Table

Configure properties of Table and Columns:

You can use Model View pane to Configure Properties

General tab, you can:

1.   Edit the name and description of the column.

2.   Add synonyms that can be used to identify the column when you are using the Q&A feature.

3.   Add a column into a folder to further organize the table structure.

4.   Hide or show the column.

Formatting tab, you can:

1. Change the data type.

2. Format the date,currency,decimal etc

Advanced tab, you can:

1. Sort by a specific column.

2. Assign a specific category to the data.

3. Summarize the data.

4. Determine if the column or table contains null values

Note: This process of formatting and configuring tables can also be done in Power Query.

Synonyms:

- This step applies specifically to Q&A.

- Users often have a variety of terms they use to refer to the same thing.

- The more sensible synonyms you add, the better your users' experience is with your report.

- Example: total sales, net sales, total net sales

Row label:

- A row label allows you to define which column (or *field*) best identifies a single row in a table

- Providing this extra metadata allows Q&A to plot a more helpful visual

- Product Table→Row Label→Category

Example:

Q&A: Show me sales by Product

Sort By Column

- In Date Table for Month column, set Sort by column as MonthNumber

Data Category:

- In Region Table Categorize Country field as "Country/Region"

Summarize by:

- You can change the summarization by different options like min, max, average etc

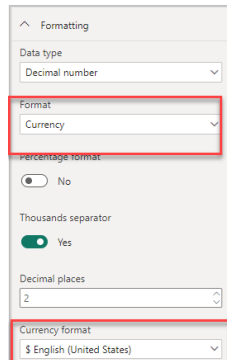**Lab 2: Configuring Column and Table Properties**

- **Provide Description For Sales Table**

- **Provide Description for SalesAmount Column in it**

- **Change the formatting options of SalesAmount and observe the change in Visual**

- **Hide all key columns in all Tables**

- **Change aggregation for Unit Price in Sales Table to Average**

- **Provide Data Categories for State and Country-region and City in Reseller Table.**

Solution:

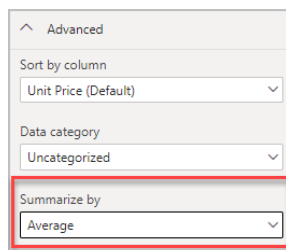1. Go to Model View→Select Sales Table→in Properties window Provide Description(This is Sales)

Note: In Report View as you hover over Table you can see the description.

2. Similarly, you can also provide description for any column.

3. Model View→Sales Table→Sales Amount column→Properties pane → set format and Currency format as shown below:
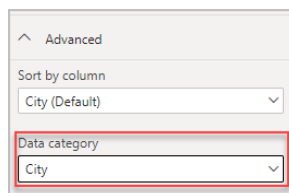
Formatting
Data type
Decimal number
Format
Currency
Percentage format
No
Thousands separator
Yes
Decimal places
2
Currency format
$ English (United States)

→Observe The change in Report View.

4. Model View→Sales Table→Hold ctrl key and select key columns(CustomerKey,DueDateKey,ProductKey,ResellerKey,SalesTerritoryKey)→Right Click→Hide in Report View

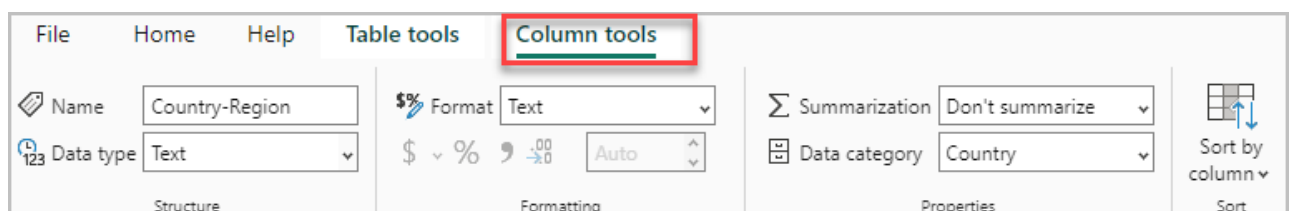5. Model View→Sales Table→Unit Price→Properties window→Advanced→Change summarization to Average

Advanced
Sort by column
Unit Price (Default)
Data category
Uncategorized
Summarize by
Average

6. Reseller→City Column→Properties window→Advanced→Data Category:City

Advanced
Sort by column
City (Default)
Data category
City

7. Similarly change Data Category for State and Country-region respectively

Note: You can also do these changes in Table View

8. Table View→Column tools Ribbon

File	Home	Help	Table tools	**Column tools**		
Name	Country-Region		Format Text	Summarization Don't summarize	Sort by column
Data type Text		$ % 9 .00 Auto	Data category Country	
Structure		Formatting	Properties	Sort

Relationship and Cardinality
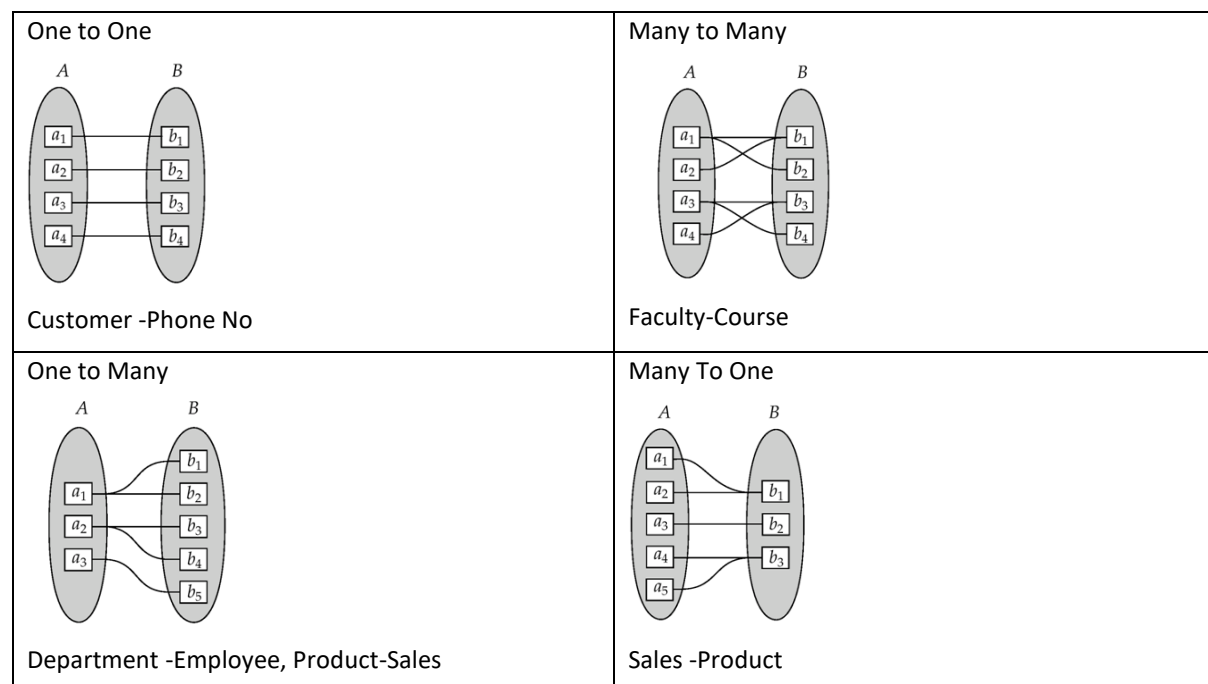
Relationship :

It is Formed by correlating rows belonging to different tables.

Cardinality and Directionality:

Relationship is expressed in terms of cardinality.

In PowerBI it plays important role in filtering data between multiple tables.

Cardinality can be :

One to One	Many to Many
Customer -Phone No	Faculty-Course
One to Many	Many To One
Department -Employee, Product-Sales	Sales -Product

- *A star icon (*) represents Many, and a 1 represents One*
- *When you click a relationship line, the related columns in either table are highlighted with a black border, for quick identification.*

Cross Filter Direction:

- Data can be filtered on one or both sides of a relationship
- *The arrow icon on the line indicates the cross filter direction of the relationship(Single/Both)*

- *There is Either **one arrow** for **Single direction** , pointing in the direction of the filter, or **two arrows** when the cross filter **direction** is set to **Both**.*

Single cross-filter direction:
- Only one table in a relationship can be used to filter the data.
- For instance, Sales can be filtered by Customer, but Customer cannot be filtered by Sales.
- You typically want these arrows to point to your fact table.

Both cross-filter directions or **bi-directional cross-filtering**:
- One table in a relationship can be used to filter the other.
- For instance, a dimension table can be filtered through the fact table, and the fact tables can be filtered through the dimension table.
- You might have lower performance when using bi-directional cross-filtering with many-to-many relationships.
- The cross-filter direction of the relationships in your dataset affect how Power BI treats the tables in visualizations in your reports.

Lab3: Understanding Cross Filter Direction

**Create Following Report**
1. How many order each customer placed
2. How many customers placed orders in a day

Note:In Model view Observer the relationship between Cusomer and Sales and Observe Cross Filter Direction

Solution:
1. Create New Report page →Select Customer|Customer and Sales|SalesOrderLineKey →Observer the result is not proper
2. Change aggregation for SalesOrderLineKey as Count
3. Visualization Section→Columns→Expand Sales OrderLineKey→Select Count
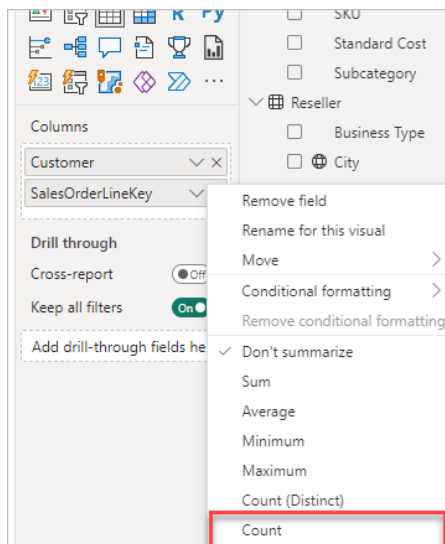
4. Create One more table visual with Sales|OrderDateKey and Customer|Customer

5. Change aggregation for Customer|Customer as Count and **observe the same customer count for all dates.**

Customer	Count of SalesOrderLineKey		OrderDateKey	Count of Customer
[Not Applicable]	60855		20170701	18485
Aaron Adams	4		20170702	18485
Aaron Alexander	1		20170703	18485
Aaron Allen	1		20170704	18485
Aaron Baker	2		20170705	18485
Aaron Bryant	5		20170706	18485
Aaron Butler	2		20170707	18485
Aaron Campbell	2		20170708	18485
Aaron Carter	2		20170709	18485
Aaron Chen	2		20170710	18485
Aaron Coleman	3		20170711	18485
Total	**121253**		**Total**	**18485**

6. Model view→Select relationship between Customer and Sales→Right Click→Properties→Change Cross Filter directions Both→Click Ok
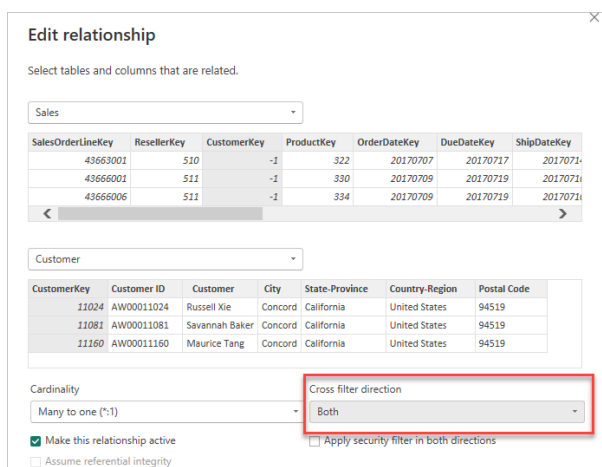
7. Observe the Change in Report View.

- Dimension tables store details about business entities, such as products or time, and are connected back to fact tables through a relationship.
- Hierarchies Organize data such that one element is ranked above other data or form natural segments in your data.

Example: Organizational Hierarchy, Date Hierarchy

Advantage of Hierarchy in PowerBI:

- Helps user understand various Levels in Data
- Enables Users to Drill up and Down in Visual
- There is performance benefit as Engine can potentially Pre-calculate the result

Types of Hierarchies in PowerBI

1. **Implicit:** Generated by Default(Date)
2. **Explicit:** Created by User

Example:

Group→Country→Region Hierarchy

Category→SubCategory→Product

Note: Hierarchies can be created based on existing columns which are part of same table.

Implicit Date Hierarchy:

Apply auto date/time in feature in PowerBI

When Enabled, Power BI Desktop creates a hidden auto date/time table for each date column, providing all of the following conditions are true:

- The table storage mode is Import
- The column data type is date or date/time
- The column isn't the "many" side of a model relationship

Navigation:

File→Options and Settings→Options→Data Load→Time Intelligence→Enable Auto Date/Time

Lab4: Apply auto date/time in PowerBI and Explore Date Hierarchy

1. Go to File→Options and Settings→Options→Observe **Data Load** option under **Global and Current File**
2. Enable Time intelligence Auto date/time at both places.
3. Data Load→Time Intelligence→Enable Auto Date/Time→OK
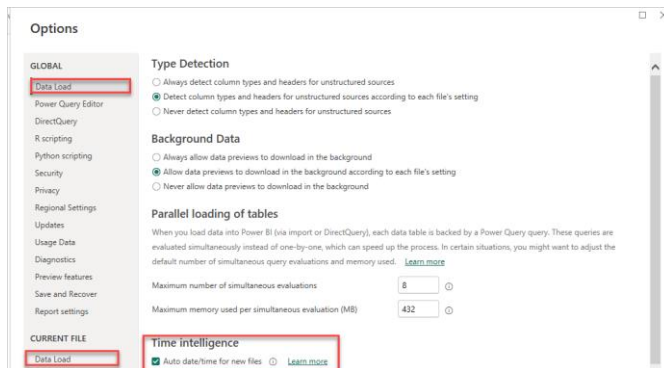
4. Observe that for each data column date hierarchy is created.

5. Create new page and select date Hierarchy and observe the table with various fields

6. Select Sales Amount from Sales Table

7. Click on Stack column chart and the visual get converted into bar chart.

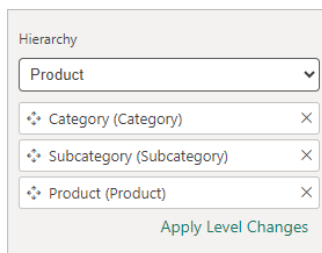8. Explore Drill down using Drill controls available on visual.
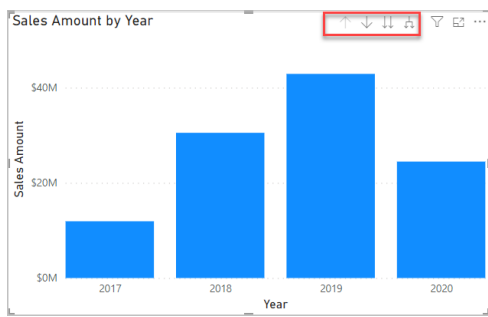
Explicit Product Hierarchy

Lab5: Create Explicit Hierarchy For Product

Model View→go to Product Table

1. Product→Category→Right Click→Create Hierarchy

2. Properties pane→Hierarchy→Select SubCategory

3. Properties pane→Hierarchy→Select Product

4. APPLY Level Changes

5. Remane Hierarchie as Products

    Properties Pane →Name: Products

6. Create Visual with Products Hierarchy and Sales Amount convert it to matrix and observe.

Note: You can create Hierarchy using Model or Report View too.

Role-playing dimensions:

- A **role-playing dimension** is a dimension that can filter related facts differently. For example, at Adventure Works, the date dimension table has three relationships to the reseller sales facts. The same dimension table can be used to filter the facts by order date, ship date, or delivery date.
- In a Power BI model, this design can be imitated by creating multiple relationships between two tables.
- But there can only be one active relationship between two Power BI model tables
- Filter propagation happens through default relationship
- To use inactive relationship as Filter define a DAX expression that uses the USERELATIONSHIP function

---

Create Date Table

Date Table can be built:

1. Using Source data date table
2. Generate Date Dimension with Power Query
3. Generate Date Dimension with DAX

Using DAX:

Navigation: Table View →Home→Calculation→New Table

1. CALENDAR:
   - The CALENDAR() function returns a contiguous range of dates based on a start and end date that are entered as arguments in the function.

     Example:

     Dates= CALENDAR (DATE (2005, 1, 1), DATE (2015, 12, 31))

2. CALENDARAUTO:
   - CALENDARAUTO() function returns a contiguous, complete range of dates that are automatically determined from your dataset.
   - The starting date is earliest date in your dataset, and the ending date is the latest date in your dataset .
   - You can pass in a single optional parameter that's the end month of the year by default it is considered as 12
   - You can give other value representing Fiscal month

     Example:

     Dates = CALENDARAUTO(6)

Note: These functions Returns a single-column table consisting of date values.

*You can create other columns in date table using DAX calculation.*

Create Other Columns Using Calculations:

**Table View→Select Dates Table→ Home→Calculations→New Column**

Year = YEAR(Dates[Date])

MonthNum = MONTH(Dates[Date])

WeekNum = WEEKNUM(Dates[Date])

DayoftheWeek = FORMAT(Dates[Date].[Day], "DDDD")

Month = FORMAT(Dates[Date], "MMM")

MonthYear = FORMAT(Dates[Date],"mmm yyyy")

Set Sort by column of Month by MonthNum

Using Power Query "M":

1. Transform→Queries→Blank space→Right Click→New Query→Blank Query→Formula bar

   =List.Dates(#date(2011,05,31), 365*10, #duration(1,0,0,0))

2. Go to the **List Tools Transform** tab on the ribbon and select **Convert > To Table**.
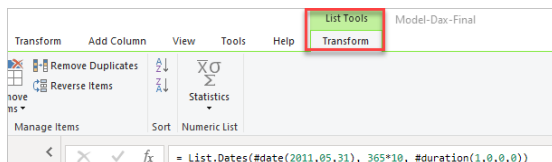
3. Rename Column as Date→Change DataType as Date

4. Select Column→Add Column→From Date & Time→Click on Date→
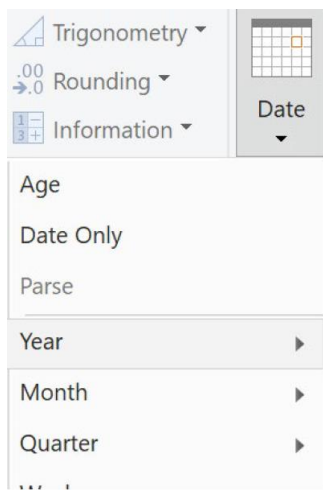
 →Similarly you can add other columns.