POWER BI – TRANSFORMATION & MODELING

Ivy Professional School

Top Ranked **Data Science Education Provider since 2007**

OUTLINE

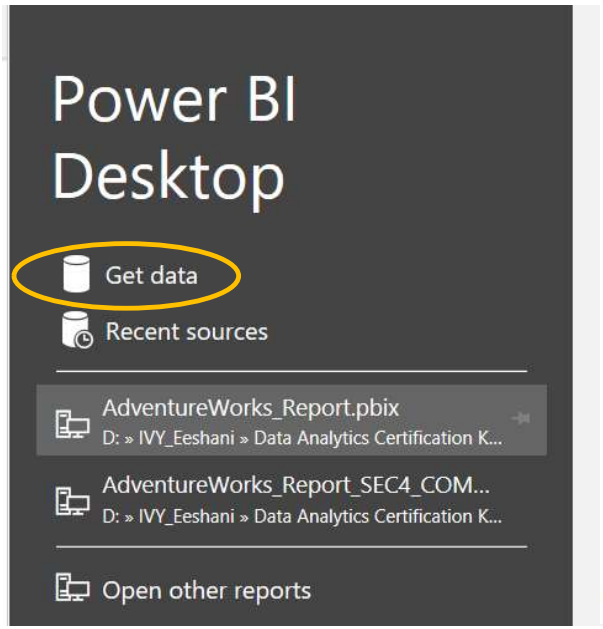
OUTLINE

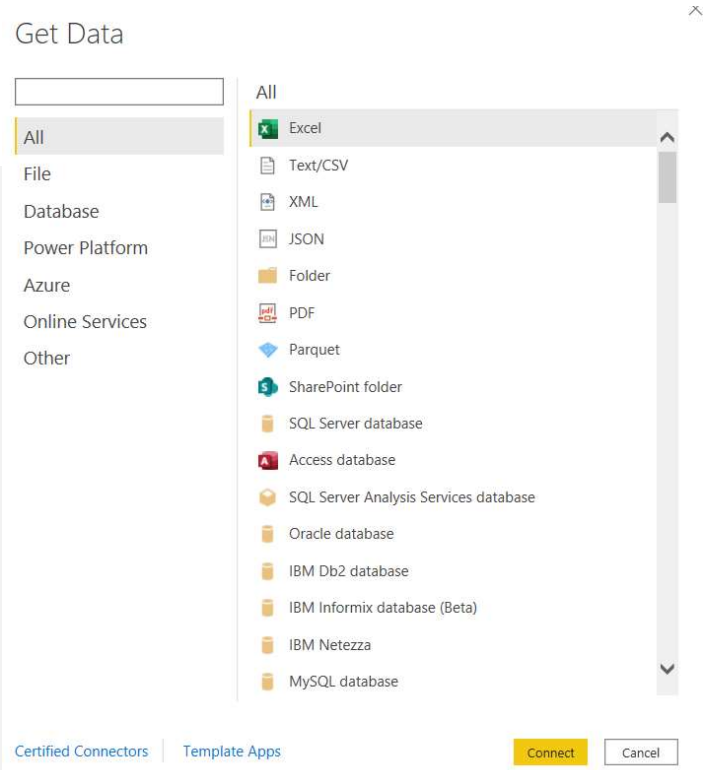
- Connecting with Data
- Data Query Editor
- Data Transformation
- Data Modelling

CONNECTING WITH DATA

Connecting with Data

- Power BI can connect with variety of Data
- Open Power BI
- Starting Screen
- Get Data
- Connect to data from file also using Get Data button in Home Tab

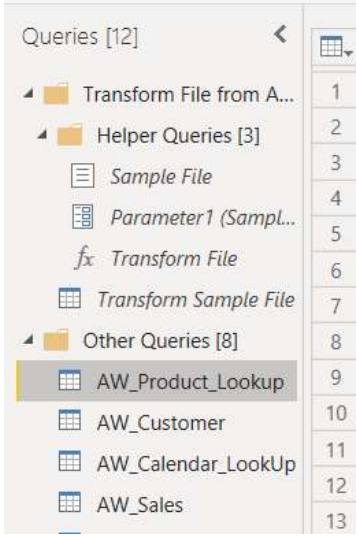




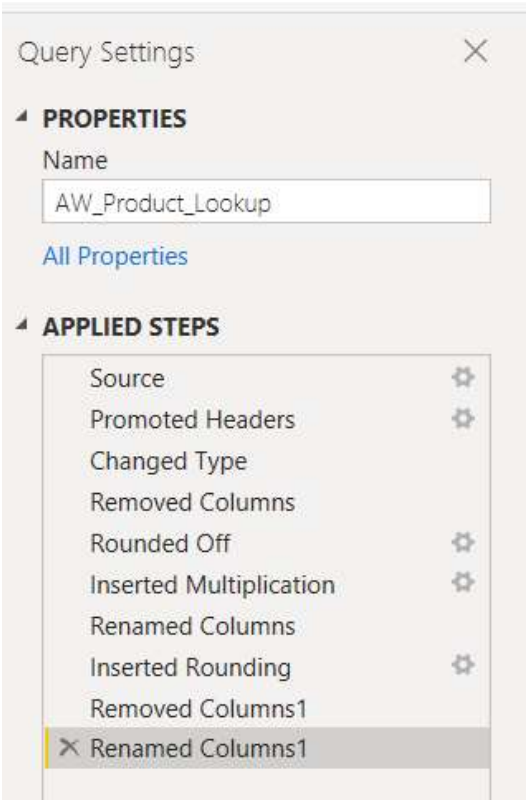
DATA QUERY EDITOR

Query Editor

- It is used for transformation of data
- Three main parts
  - Left – all queries
  - Middle – selected table
  - Right – Properties of the table and Applied Steps

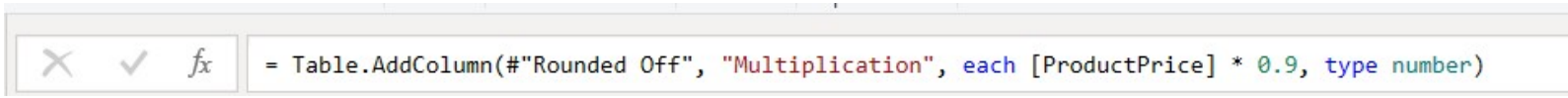


	ProductKey	ProductSubcategoryKey	ProductSKU
1	214	31	HL-U509-R
2	215	31	HL-U509
3	218	23	SO-B909-M
4	219	23	SO-B909-L
5	220	31	HL-U509-B
6	223	19	CA-1098
7	226	21	LJ-0192-S
8	229	21	LJ-0192-M
9	232	21	LJ-0192-L



Applied Steps

- Like Macros of Excel
- Automates the process
- Runs through the same set of steps again when data is refreshed
- We can delete steps and change the sequence
- Mcode shows up in Formula bar is used
- It is Power BI scripting language



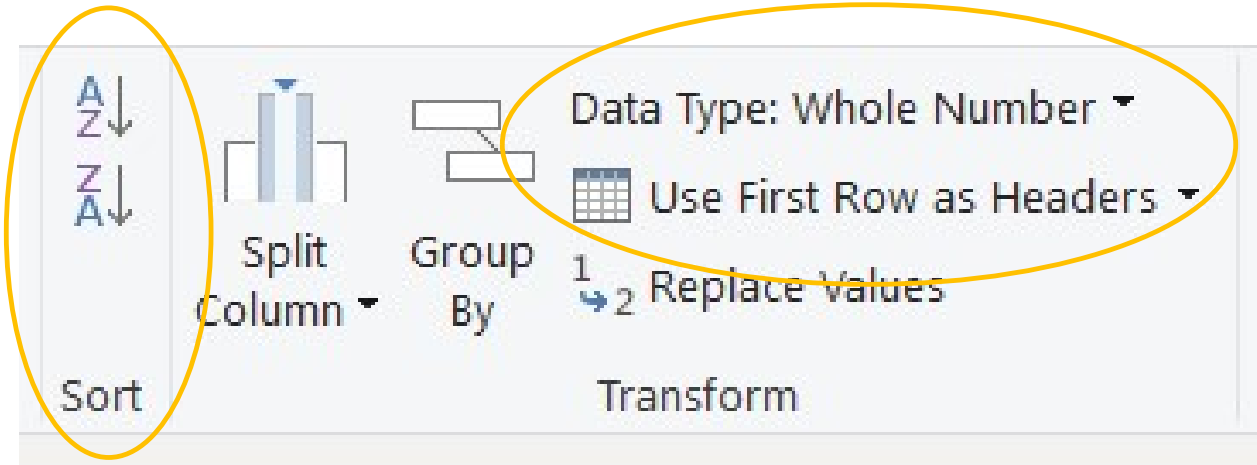
DATA TRANSFORMATION

Column and Row

- Remove unwanted columns
  - Columns not required for analysis
  - Helps to reduce the processing speed
- Remove unwanted row
  - Footers
  - Notes
  - Headings

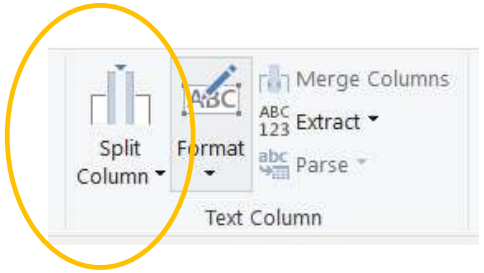
Basic Table Transformation

- Check data types of each field
  - To change data type –
    - Right click → Change Type → Select new data type
    - Or, Transform Tab → Data Type Tool on top
- Sorting columns
- Promote header rows



Text Tools #1

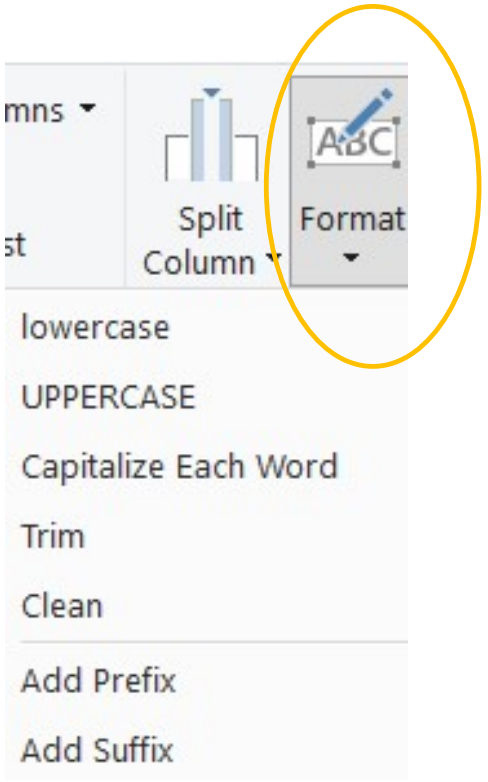
- Under Transformation Tab
- Split



i. Delimiter – based on any character
ii. Position – splits into multiple columns with specified number of char. Mention the position starting from 0 and in asc order like Eeshani → 0,5 →Eesha / ni
iii. Number of character
iv. Upper to Lower Case – Split from where Upper case changes to lower case. Splits into two columns
v. Lower to upper – similar to upper to lower
vi. Digit to Non-digit – splits from where the numbers end and alpha starts. First column will be number and second non digit
vii. Non-digit to Digit – opposite of Digit to Non-Digit

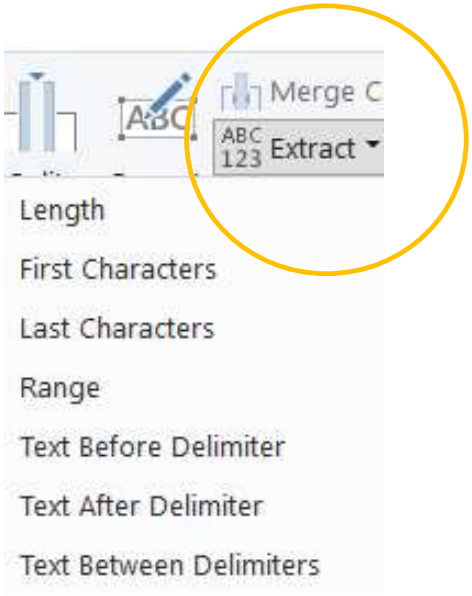
Text Tools #2

- Format
  - Changes cases – lower, upper, Capitalize Each word
  - Trim – remove leading and trailing spaces
  - Clean – removes non-printable characters
  - Add – Prefix and Suffix



Text Tools #3

- Extracts
  - Data extraction from each row of table
  - Len
  - First Character etc
- Merge
  - Creates new column by concatenation
  - Select columns to merge then Merge will activate



Exercise

Customer file

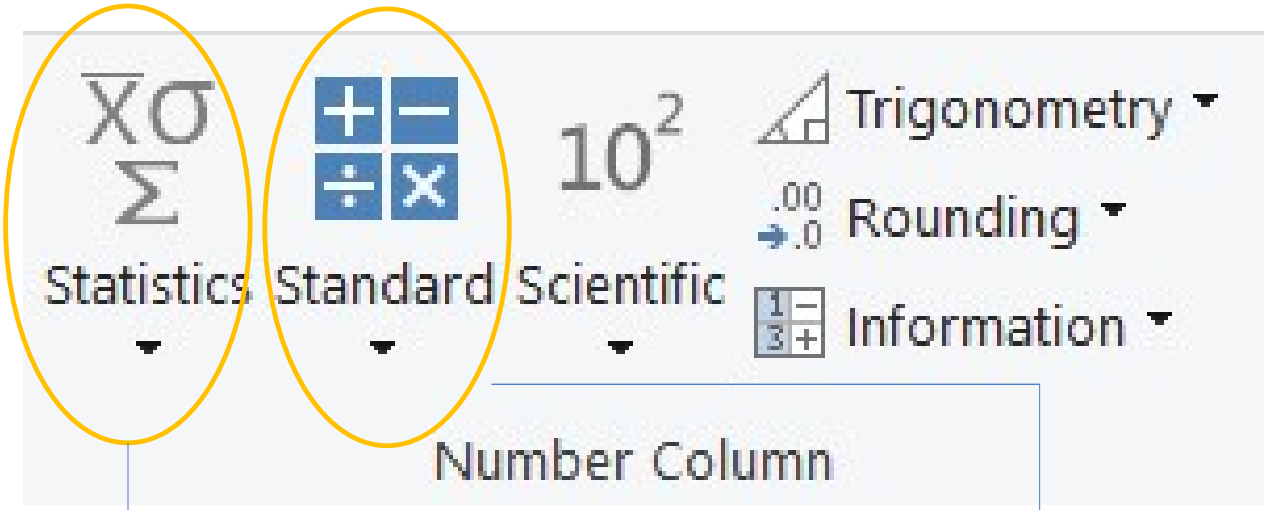
a. Convert the name into proper cases

b. Create a new column with full name

c. Create a new columns with user name and domain name from the email ID

d. In Domainname col replace '–' with space (use Replace Value Tool)

Number Tools



- Statistics is available in Transform Tab and Not Add Column
- It aggregated values in column and returns one value only
- Not applied on row level but on column level
- Used to get quick summary of data

- Standard is available in Transform Tab and Add Column
- Used to Add, Subtract, Divide etc on row level
- Used to get modify data or create new columns

Date Tools

- Date and time tools will be active when data type of selected column is date/time
- Used to extract
  - Age based on the date and today's date
  - Various components of dates like Year, Quarter, Month etc.
- Earliest and Latest will return one single value
  - It is Transformation Tab only
- Start of the week can be defined using the Mcode

= Table.AddColumn(#"Inserted Day Name", "Start of Week", each Date.StartOfWeek([Date]**, 1**), type date). Add ,1 since 1 represents Monday. Or Table.AddColumn(#"Inserted Day Name", "Start of Week", each Date.StartOfWeek([Date]**, Day.Monday**), type date). Add

Exercise

- Calculate the Age of the Customer

Index and Conditional Column

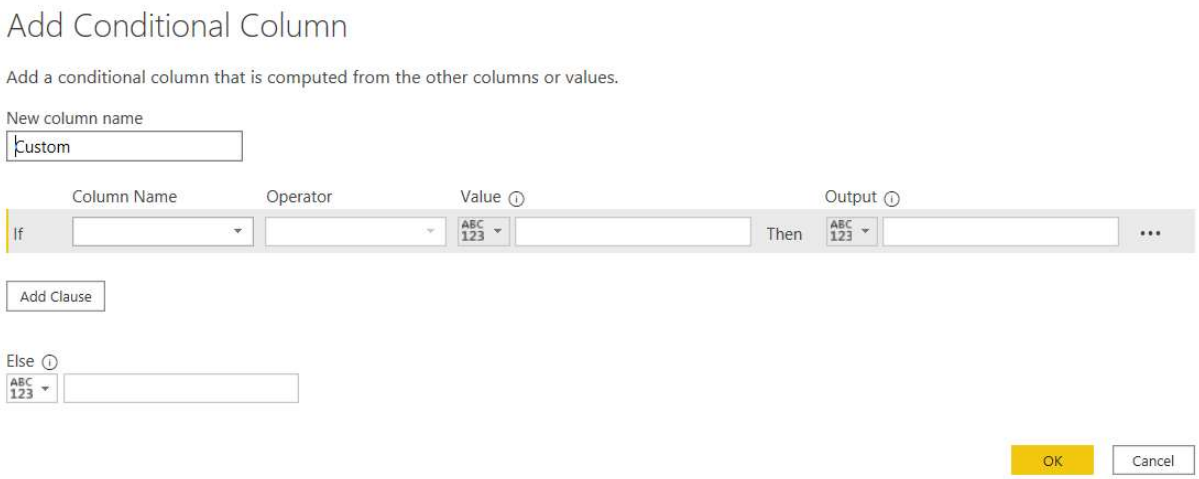
1. Index Column –

a. It is used to create sequential number column. Used for creating the Unique key for the tables which are used in the relationships.

2. Conditional Columns –

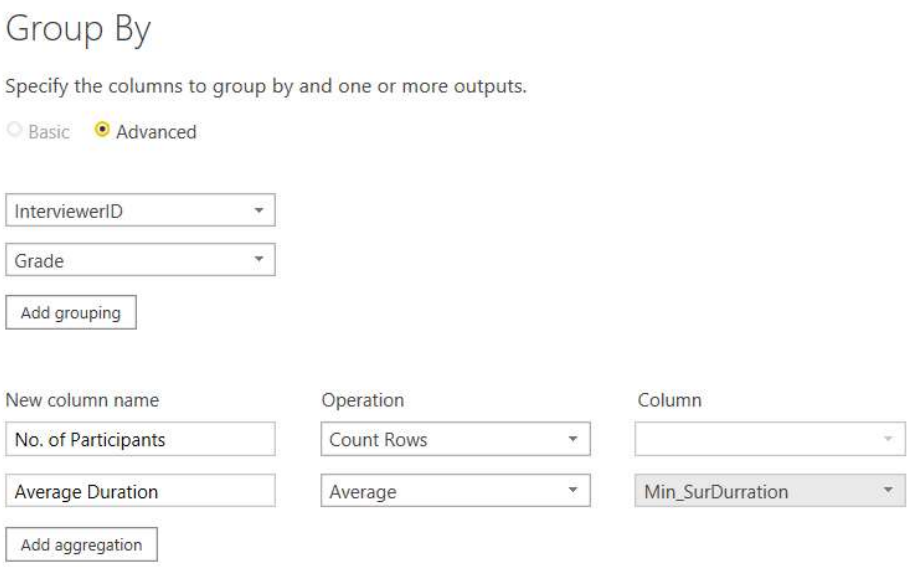
a. Helps to create new columns based on logical rules and conditions (If/Then Statement)

b. Goes through each row



Group By

- Available only in Transformation and Not Add Column
- Used to create aggregated data based on a categorical field
- In order to add multiple columns in the new table use Advanced Grouping



Pivoting and Unpivoting

- Used when data structure needs to be modified for better analysis
- Pivoting – transforms row data into column
- Unpivoting – transforms column data into row
- Pivoting and Unpivoting recognizes unique values unlike Transpose

	Age Group	Month	Male	Fe
1	15-20 y	01-01-2017	2235	
2	15-20 y	01-02-2017	3400	
3	15-20 y	01-03-2017	4970	
4	15-20 y	01-04-2017	4500	
5	21-25 y	01-01-2017	4039	
6	21-25 y	01-02-2017	2792	
7	21-25 y	01-03-2017	4765	
8	21-25 y	01-04-2017	3127	

	Age Group	Month	Attribute	Value
1	15-20 y	01-01-2017	Male	2235
2	15-20 y	01-01-2017	Female	2022
3	15-20 y	01-02-2017	Male	3400
4	15-20 y	01-02-2017	Female	4628
5	15-20 y	01-03-2017	Male	4970
6	15-20 y	01-03-2017	Female	1003
7	15-20 y	01-04-2017	Male	4500
8	15-20 y	01-04-2017	Female	2363
9	21-25 y	01-01-2017	Male	4039
10	21-25 y	01-01-2017	Female	3501
11	21-25 y	01-02-2017	Male	2792
12	21-25 y	01-02-2017	Female	3852
13	21-25 y	01-03-2017	Male	4765
14	21-25 y	01-03-2017	Female	4405
15	21-25 y	01-04-2017	Male	3127
16	21-25 y	01-04-2017	Female	2492

Exercise

AdventureWorks_Sales_2015-

a. Create a Discount Column – 90% of the Price

b. Round the columns with Price and cost

c. Find the total number of Unique products

d. Find the Average Sales and Total Sales Value

Merging Queries

- It works like Vlookup in excel where we are getting information from one table to another using a common column.
- It add columns to an existing table
- But it is better to create relationships and let the tables be separate
- This process may lead to creation of a lot of redundant data

Appending Queries

- It adds rows to the existing table
- It is like Union of tables
- The tables we want to append should have the same structure and same data type
- If we have a folder in which all the files which we need to append are kept, we can choose Folder option from GET Data option. Using the Folder option the advantage is that if any file is added or removed it gets updated when the power BI file is refreshed.
- The tables used in the queries – Merging or Appending can’t be deleted as they are being used.

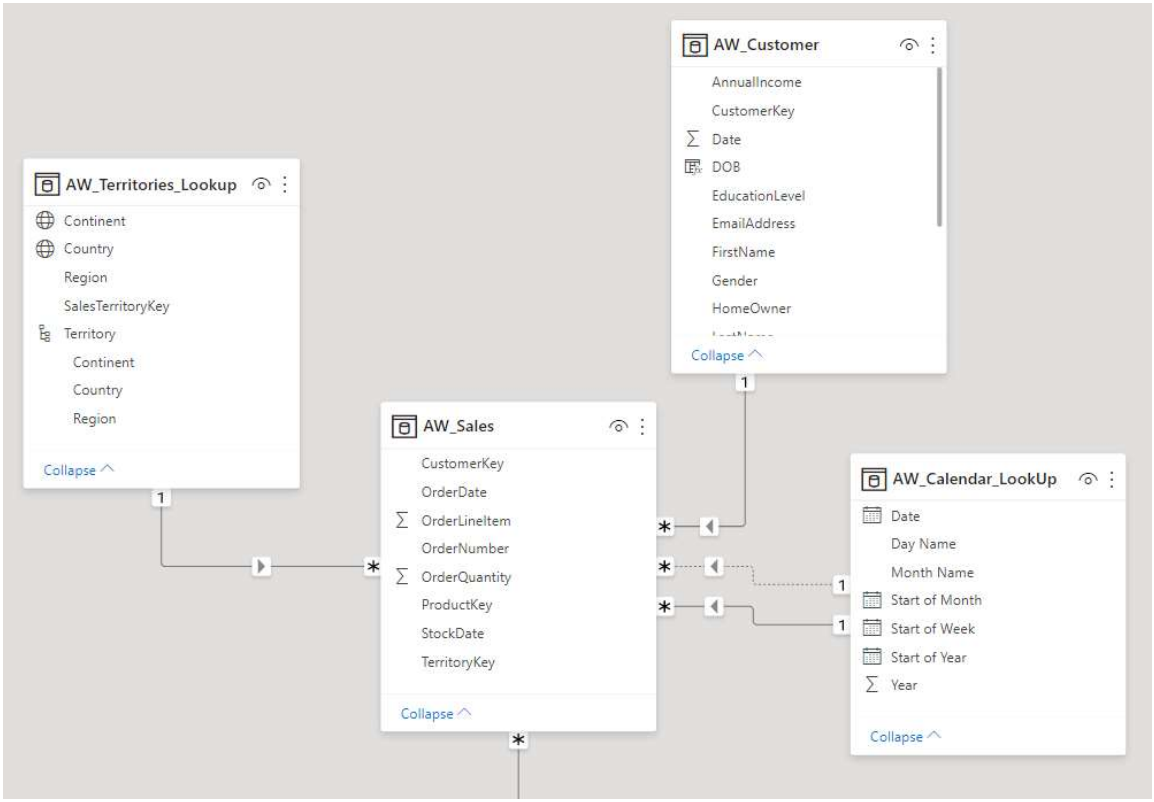
Exercise

- Create a combined table for all the sales files

DATA MODELLING

Data Models

- Tables which are connected to each other via relationships based on common fields.



Fact and Data Table

- Fact Table/Lookup Tables
  - Unique Values
- Data Tables
  - Multiple values

OrderDate	StockDate	OrderNumber	ProductKey	CustomerKey	TerritoryKey	OrderLineItem	OrderQuantity
05 July 2015	03 June 2002	SO46718	360	12570	9	1	1
07 July 2015	22 April 2002	SO46736	360	12341	9	1	1
12 July 2015	05 May 2002	SO46776	360	12356	9	1	1
16 July 2015	22 June 2002	SO46808	360	12347	9	1	1
18 July 2015	11 May 2002	SO46826	360	12575	9	1	1
01 August 2015	21 April 2002	SO47075	360	12685	9	1	1
04 August 2015	01 May 2002	SO47098	360	12667	9	1	1
10 August 2015	21 April 2002	SO47149	360	12669	9	1	1
17 August 2015	04 June 2002	SO47212	360	12580	9	1	1
26 August 2015	29 June 2002	SO47302	360	12670	9	1	1
29 August 2015	12 August 2002	SO47328	360	12681	9	1	1
31 August 2015	13 August 2002	SO47346	360	12585	9	1	1
02 October 2015	12 June 2002	SO47744	360	12989	9	1	1
02 October 2015	28 July 2002	SO47745	360	12998	9	1	1
03 October 2015	22 August 2002	SO47753	360	13020	9	1	1

Database Normalization

- Normalization is the process of organizing the tables and columns in a relational database to reduce redundancy and preserve data integrity.
- Tables should have distinct and specific data

Exercise

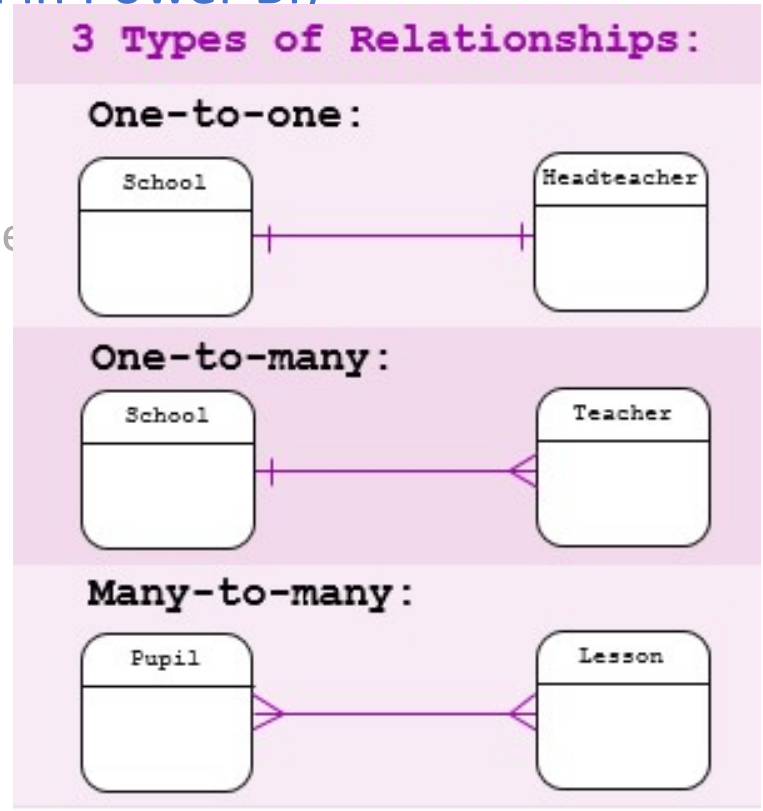
- Create tables from the Database shown following the rules of Normalization
  - Identify them as Fact or Data Table

Primary and Foreign Key

- Primary Keys are unique and exist in Fact/Lookup tables
- Foreign keys may be duplicate and they exist in Data tables
- Primary and Foreign Keys are used to create relationships between the tables

Types of Relations

- One to many (used in Power BI)
- Many to many
- One to one
  - It’s better to Merge



Creating Relationship

- Two Ways
  - Click and drag to another table on the key match
  - manage relationship button. More manual.
- Delete Relationship
- Edit Relationship
- Active vs Inactive Relationship
  - Multiple relationships from one table to another table are not allowed
  - One relationship can be active and shown by solid line
  - Another relationship will be inactive and shown by dotted line

Important - Relationships

- NO relationship created between two Data tables
  - It results in Many to Many relationship
- They are connected through Fact/Lookup Tables

Filter Flow

- When we create relationships between two tables we can see an arrow on the relationship lines. This indicates Filter Flow.
- Filter flow is always from one side to many side of the relationship.
- When we filter the table, the filter context is passed to all the downstream tables.
- Filters cannot flow upstream.

The END