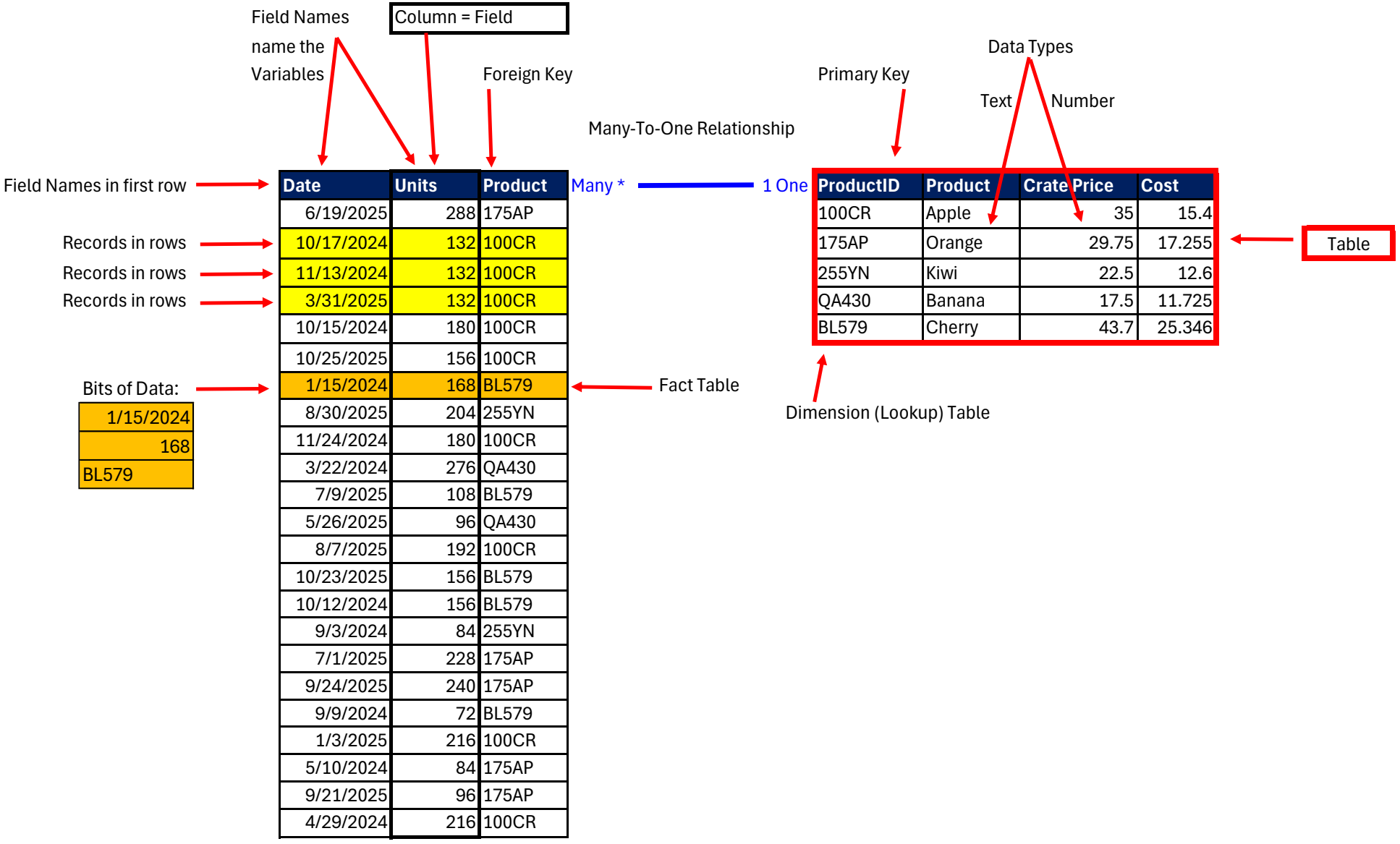
No.	Topics for Week 1 & 2:
1	Data Analysis Terminology
2	Worksheet Formulas
3	PivotTables
4	Dynamic Spilled Arrays
5	Power Query
6	Power Pivot
7	Power BI Desktop
8	Power BI Online Service
9	Dataflow (Online Power Query)


Field Names
name the
Variables
Column = Field
Foreign Key
Primary Key
Data Types
Text
Number
Many-To-One Relationship
Field Names in first row
Records in rows
Records in rows
Records in rows
Bits of Data:
1/15/2024
168
BL579
Many *
1 One
Fact Table
Dimension (Lookup) Table
Table
Date
Units
Product
6/19/2025 288 175AP
10/17/2024 132 100CR
11/13/2024 132 100CR
3/31/2025 132 100CR
10/15/2024 180 100CR
10/25/2025 156 100CR
1/15/2024 168 BL579
8/30/2025 204 255YN
11/24/2024 180 100CR
3/22/2024 276 QA430
7/9/2025 108 BL579
5/26/2025 96 QA430
8/7/2025 192 100CR
10/23/2025 156 BL579
10/12/2024 156 BL579
9/3/2024 84 255YN
7/1/2025 228 175AP
9/24/2025 240 175AP
9/9/2024 72 BL579
1/3/2025 216 100CR
5/10/2024 84 175AP
9/21/2025 96 175AP
4/29/2024 216 100CR
ProductID
Product
Crate Price
Cost
100CR Apple 35 15.4
175AP Orange 29.75 17.255
255YN Kiwi 22.5 12.6
QA430 Banana 17.5 11.725
BL579 Cherry 43.7 25.346

Invoice Total Sales = Less Detail, Bigger Grain

Line Items Sales = More Detail, Smaller Grain

Date	Invoice Number	Sales RepID	Shipping Costs($)	Invoice Discount($)
1/1/2017	125447	9	98.7	144.18
1/2/2017	125448	28	26.25	73.06
1/3/2017	125450	4	207.55	437.62
1/4/2017	125451	15	262.15	542.26
1/4/2017	125452	23	159.25	381.63

Invoice Number	Product ID	Quantity	Unit Price($)
125447	LS-900	21	22.36
125447	TC-500	88	14.97
125447	OK-800	35	12.32
125448	DQ-100	53	27.57
125450	SC-1100	25	48.75
125450	TC-500	34	16.22
125450	TS-300	200	13.03
125451	AC-1000	223	11.68
125451	LS-900	224	12.58
125452	TM-600	38	18.82
125452	IY-700	238	13.03

Year Grain is Bigger than Month Grain

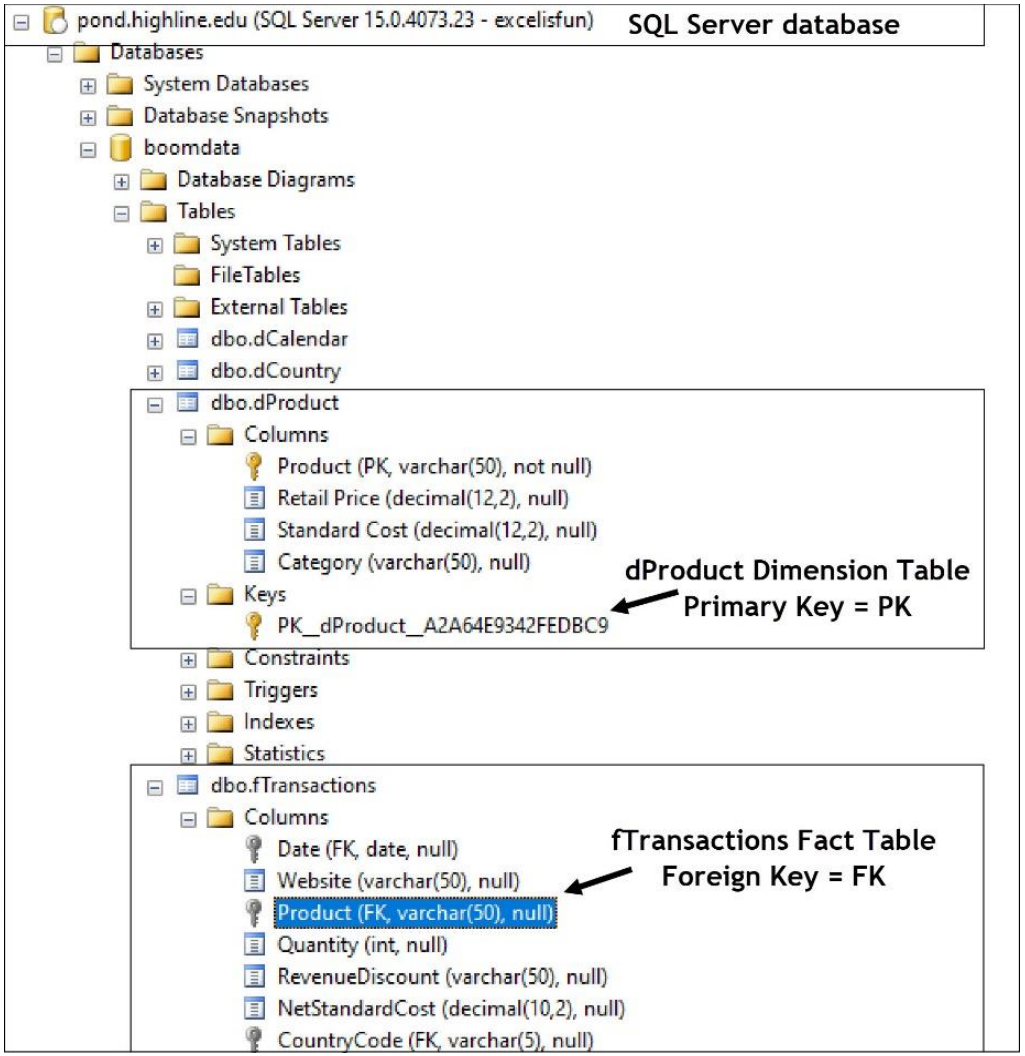
Year Grain has Less Detail than Month Grain

Year	Sales($)
2023	78,139
2024	58,066

Year	Month	Sales($)
2023	Jan	7,684
2023	Feb	8,255
2023	Mar	3,286
2023	Apr	8,572
2023	May	2,829
2023	Jun	7,296
2023	Jul	5,731
2023	Aug	5,610
2023	Sep	8,218
2023	Oct	6,507
2023	Nov	12,862
2023	Dec	1,289
2024	Jan	12,683
2024	Feb	4,900
2024	Mar	1,999
2024	Apr	1,720
2024	May	1,343
2024	Jun	2,840
2024	Jul	7,475
2024	Aug	10,813
2024	Sep	1,249
2024	Oct	580
2024	Nov	9,355
2024	Dec	3,109

SQL Server Database
Relational Database

Much of the data in the world is stored in SQL databases



Columnar database in the Power Pivot and Power BI Data Model and Sematic Model compresses data into a smaller more efficient structure than just storing rows of data.

Date	ProductID	SalesRepID	Units
3/19/2021	2	4	80
4/8/2021	2	4	5
4/12/2021	2	4	88
4/12/2021	2	4	70
4/22/2021	2	4	6
5/12/2021	2	4	5
5/24/2021	2	4	1
6/11/2021	2	4	92
6/11/2021	2	4	91
6/17/2021	2	4	209
7/5/2021	2	4	91
7/18/2021	2	4	66
7/20/2021	2	4	110
7/24/2021	2	4	4
8/6/2021	2	4	94
8/18/2021	2	4	2
9/10/2021	2	4	57
10/2/2021	2	4	4
10/2/2021	2	4	1
10/7/2021	2	4	86
10/7/2021	2	4	90

Records in original table:	606
# Columns:	4
Total cells with data:	2424

Total cells in columnar database:	621

Count
426

Date
3/19/2021
4/8/2021
4/12/2021
4/22/2021
5/12/2021
5/24/2021
6/11/2021
6/17/2021
7/5/2021
7/18/2021
7/20/2021

Count
4

ProductID
2
3
4
1

Count
4

SalesRepID
4
2
3
1

Count
187

Units
80
5
88
70
6
1
92
91
209
66
110

Source data	The original location of the data, like in a text file, an Excel file of a database.
Data destination	The location where the data is loaded, such as in an Excel or a Power BI Desktop file or an online source like a Power BI workspace.
On-premises file path	A hard coded source data file path in the data destination, such as an Excel or Power BI Desktop file.
	On-premises file paths can cause errors when the data destination file is moved and the connection to the source data is lost.
Online source data	Online source data can solve the problem of On-premises file and folder paths.
	Web sites, SQL Server databases and Power BI Online are examples of online sources that stay connected to the source data when the data destination file is moved.
Delimiter	Is a character that separated bits of data, such as a comma, tab and other characters.
Structure Or Schema	The rules or structure for tables, data files and databases.

Source data in files.

Text Files:

Csv (Comma Separated Values)
Table Structure/Schema: field names & data are separated by commas (delimiter).
Systems understand that the data is stored with a table structure.

```
Date,Units,Product,Customer,SalesRep¶
6/19/25,288,175AP,SFM,266SB¶
10/17/24,132,100CR,PCC,513CM¶
```

Txt (Tab Separated Values)
Table Structure/Schema: field names & data are separated by Tabs (delimiter).
Systems understand that the data is stored with a table structure.

```
Date→Units→Product → Customer → SalesRep¶
6/19/2025 → 288 → 175AP→SFM → 266SB¶
10/17/2024→132 → 100CR→PCC → 513CM¶
```

Xml (eXtensible Markup Language)

Table Structure/Schema:

A "Sales" table container contains fields.

Transaction record containers contains fields.

Various field containers contain data.

Systems understand that the data is stored with a table structure.

```
<?xml version="1.0" encoding="UTF-8"?>
<Sales>
  <Transaction>
    <Date>6/19/25</Date>
    <Units>288</Units>
    <Product>175AP</Product>
    <Customer>SFM</Customer>
    <SalesRep>266SB</SalesRep>
  </Transaction>
  <Transaction>
    <Date>10/17/24</Date>
    <Units>132</Units>
    <Product>100CR</Product>
    <Customer>PCC</Customer>
    <SalesRep>513CM</SalesRep>
  </Transaction>
```

Json (JavaScript Object Notation)

Record Structure/Schema:

All records are housed in square brackets [ ].

Individual records are housed in curly brackets { }.

Delimiter is colon :.

There is no table schema. Json files store records, not tables.

Systems understand that the data is stored with a record structure.

After you import records, you must convert records into a table.

```
[
  {
    "Date": "6/19/2025",
    "Units": 288,
    "Product": "175AP",
    "Customer": "SFM",
    "SalesRep": "266SB"
  },
  {
    "Date": "10/17/2024",
    "Units": 132,
    "Product": "100CR",
    "Customer": "PCC",
    "SalesRep": "513CM"
  },
```

Excel files

Worksheets

Record Structure/Schema:

Data is stored in worksheet and is understood by systems to be records of data.

After you import records, you must convert records into a table.

	A	B	C	D	E
1	**Date**	**Units**	**Product**	**Customer**	**SalesRep**
2	6/19/2025	288	175AP	SFM	266SB
3	10/17/2024	132	100CR	PCC	513CM

Excel Tables

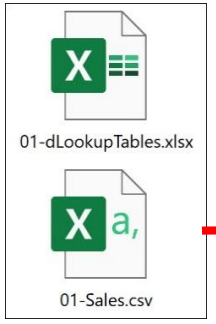
Table Structure/Schema:

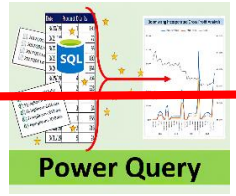
Data is stored in an Excel Table object and is understood by systems to be a table of data.

Data is stored in an Excel Table object and is understood by systems to be a table of data.

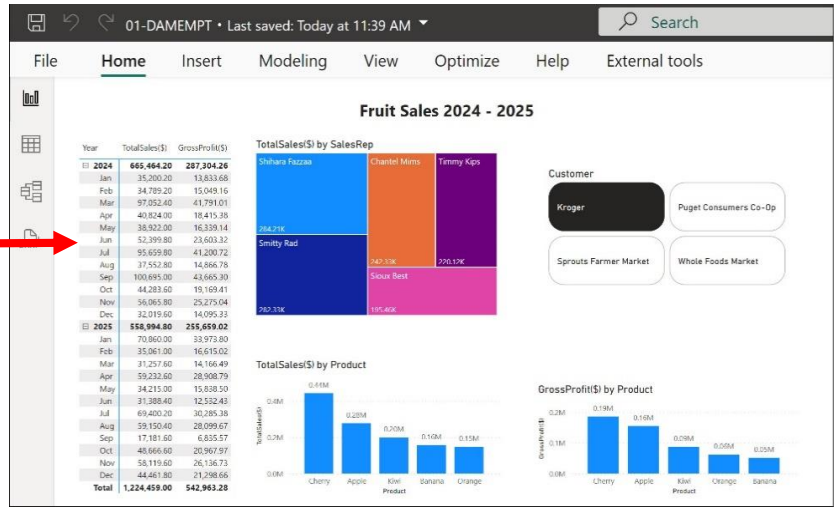
	A	B	C	D	E
1					
2		ProductID	Product	CratePrice	CrateCost
3		100CR	Apple	35	15.4
4		175AP	Orange	29.75	17.255
5		255YN	Kiwi	22.5	12.6
6		QA430	Banana	17.5	11.725
7		BL579	Cherry	43.7	25.346
8					

Source data from on-premises files.





Data destination is this Power BI Desktop file.



On-Premises file path:

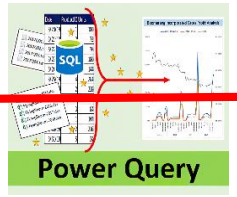
Path is broken if destination file is moved.

(Easy to fix, but annoying)

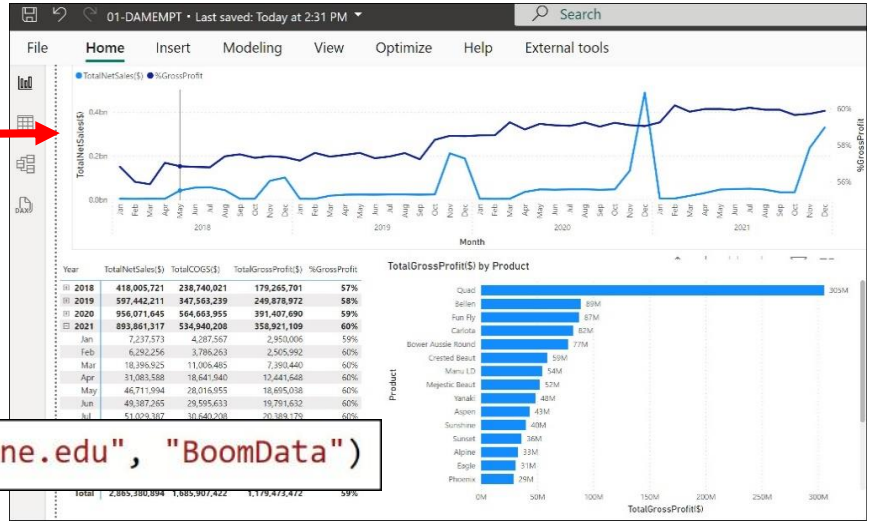
```
= Csv.Document(
    File.Contents("E:\01-Sales.csv"),
    [Delimiter=","])
```

```
= Excel.Workbook(
    File.Contents("E:\01-dLookupTables.xlsx"))
```

Online Source like SQL Database:



Data destination is this Power BI Desktop file.



Online SQL Server database

Online data source

Path is NOT broken if destination file is moved.

```
= Sql.Database("pond.highline.edu", "BoomData")
```

When you hard coded a file path into a query, it means that if you move that source data file to a different location, or you e-mail the file that contains the query to a colleague, the connection to the source data is broken and you will receive the following error: **Data Source Not Found**. When you hard code a file or folder path into a query, the path is called an **on-premises path**. As you can imagine, on-premises paths cause a lot of trouble if you are sharing files or moving files around. An on-line data source, like an SQL server database, Dataflow or Power BI do not have this problem because the data is stored online in a location that does not move. Multiple credentialed people can have access to "a single source of truth" where there are no on-premises paths and no conflicts with multiple versions of the same file. Nevertheless, not all data is stored online, and on-premises paths are common. The good news is that if you know where the source data file is, it is easy to redirect the query to the new location. There are at least three ways to change the on-premises path:

- In the Source step for almost any query, you can edit the on-premises path in the formula bar.
- You can click the gear icon in the Source query step to open the source data dialog box. Many data sources such as Csv, Excel, Sql Databases, Web sites and more allow you to use the gear icon in the source query step to edit the connection details.
- If you have used the same file or folder in multiple queries, it is most efficient to edit the path universally in the Data Source Settings dialog box. There are multiple ways to open this dialog box in Excel and Power BI. If you are in the Power Query Editor:
    - In the Excel Power Query Editor, in the Home tab, Data Sources group, click the Data source settings button.
    - In the Power BI Desktop Power Query Editor, in the Home tab, Data source group, click the Transform data dropdown and then click Data source settings.
    - In the Dataflow Power Query Editor, in the Home tab, Data sources group, click the Manage connections button.

Data Analysis (Data Analytics, Analytics, Business Intelligence, Data Science):

Define:

Converts draw data into useful, actionable information to gain insight and make decisions.

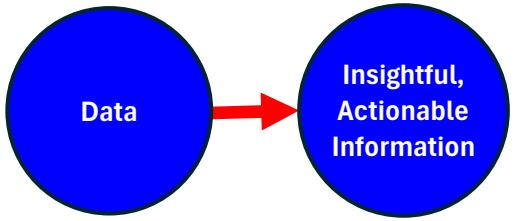
Information can be in the form of: reports, visuals, dashboards, and other forms.

Data analysis allows you to make data-driven decisions, which tend to be more accurate & help to achieve goals more consistently.

Business Intelligence:

Same definition, but the process is performed within the context of business data and business decision making.

High Level Definition:



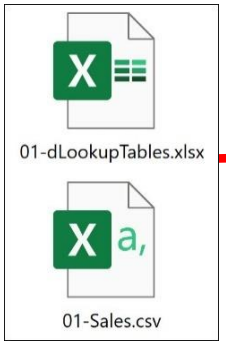
Data analysis process:

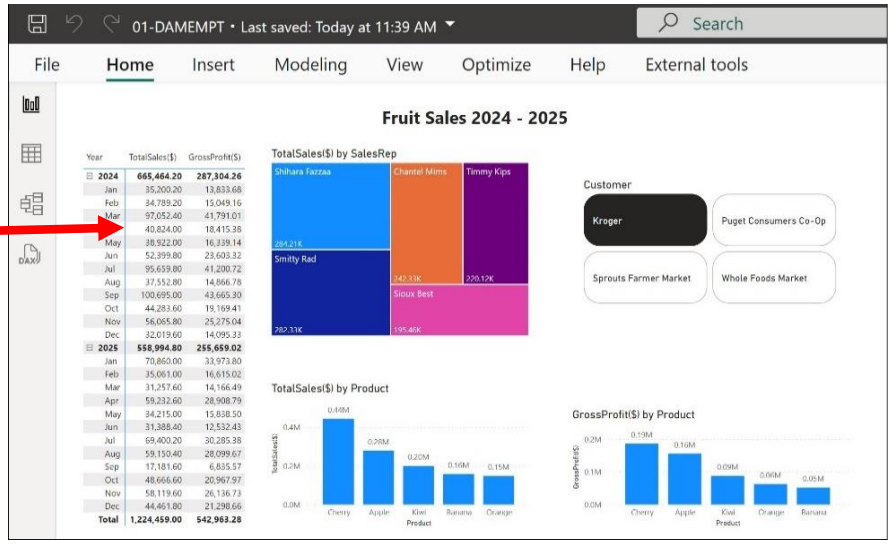
1. Determine what questions need answers & what decisions need to be made. Everything else in the process is dictated by these questions and decisions.
2. Where is the data? How much data? What is the structure of the data?
3. Which MS tool to use? (Almost always starts with Power Query).
4. Clean, transform and shape the data into a table or model that is best suited to answer questions and make decisions,
5. Build final model with measures, metrics, relationships, and other features.
6. Create useful information: reports, visuals and dashboards.
7. Refresh when new data arrives.
8. Change and update model as necessary.

Examples:

Business context:

Data	Useful information: metrics to see how customers sales perform over months, sales reps and products.





Sports contexts:

Data	Decide who becomes NBA 2024 MVP.

Rank PPG	PLAYER	TEAM	GP	MIN	Ave PPG	FGM	FGA	FG%
1	Luka Doncic	DAL	63	37.4	33.9	11.5	23.7	48.7
2	Giannis Antetokoun	MIL	67	35.1	30.6	11.5	18.7	61.4
3	Shai Gilgeous-Alexa	OKC	70	34.4	30.4	10.8	20	54
4	Kevin Durant	PHX	66	37.2	27.6	10.1	19.2	52.9
5	Jalen Brunson	NYK	67	35	27.4	9.8	20.5	47.7
6	Jayson Tatum	BOS	67	35.8	27.3	9.2	19.5	47.3
7	Devin Booker	PHX	59	35.8	27.2	9.6	19.5	49.2
8	Stephen Curry	GSW	66	32.7	26.5	8.8	19.6	44.7
9	De'Aaron Fox	SAC	64	35.6	26.5	9.7	20.8	46.6
50	Collin Sexton	UTA	73	26.3	18.6	6.5	13.2	49.4

Educational content

Data	**Decision:** Which students get scholarships? History or Sociology majors are eligible.

Oracle Database

Student	Start Date	Major	Credits	GPA	Eligible?
Coats, Saharra	9/29/2020	Sociology	45	1.7	TRUE
Emmons, Christi	7/14/2018	Accounting	135	2.3	FALSE
Lear, Vania	9/3/2020	Chemistry	45	3	FALSE
Washington, T	11/21/2019	History	90	3.1	TRUE
Mohamed, Abdi	1/28/2021	Business	23	1.6	FALSE
Nga, Luong	7/7/2020	Physics	45	2.4	FALSE
Mims, Chantel	4/12/2020	History	70	4	TRUE
Rouse, Sioux	6/30/2020	Chemistry	40	2.4	FALSE
Simone, Alanna	8/2/2019	Physics	60	3.5	FALSE
Thornburg, Tyrone	12/27/2019	Sociology	75	3.9	TRUE

Major:
History
Sociology

Personal Budgeting context:

Data keep in Excel worksheet:	**Insight:**	What are the top 5 expenses so far this year?

Date	Expenses	Amount
1/3/24	Weekend Trip	182.1
1/3/24	Medical	409.54
1/8/24	Utilities	136.37
1/10/24	Ikea	259.63
1/16/24	Car Insurance	102.5
1/19/24	Movies	43.69
1/23/24	Groceries	35.2
1/26/24	Garden	12.75
2/1/24	Gas	40
2/9/24	Utilities	92.31
2/19/24	Gas	48.2
2/27/24	Gas	55.8
2/27/24	Car Insurance	102.5
3/2/24	Groceries	102.52
3/3/24	Gas	50

Top	5

Date	Expenses	Amount
1/3/2024	Medical	409.54
3/28/2024	B-day for for Mom	378.02
1/10/2024	Ikea	259.63
1/3/2024	Weekend Trip	182.1
4/1/2024	Groceries	140.3

Data modeling

Converting source data into a data structure that allows you to create the useful information.

Tools to perform data modeling

Worksheet formulas
Power Query M Code formulas
Data Model DAX formulas
Others too

Query

A question that we ask of the raw data, or an action taken to shape data, like:
Import Csv file, append tables, or group line item sales to calculate invoice total sales.

Data modeling tasks:

Cleaning data

ISO Date	Into =>	**Proper Date**	Formula:
20240522		5/22/2024	=TEXT(B20,"0000-00-00")+0
			Or other methods in Power Query

Transforming data

Really Bad Data Into => **A Proper table that can be used for analysis**

Excel Table = **TvAppData**

Data
App: Freevee
The best free streaming service overall
SPECIFICATIONS
Live TV: Yes
Originals: Yes
Registration required: Yes
Supported devices: Amazon Fire TV, Fire TV
TODAY'S BEST DEALS
Go to Freevee
REASONS TO BUY
+
The roku channel home screen
(Image credit: Tom's Guide)
App: Roku Channel

App	Live TV	Originals	Registration required
Freevee	Yes	Yes	Yes
Roku Channel	Yes	Yes	No
Pluto TV	Yes	No	No
Tubi	No	Yes	No
Crackle	No	Yes	No
Sling Freestream	Yes	No	No
Vudu	No	No	Yes
Xumo	Yes	No	No
Peacock	Yes	Yes	No

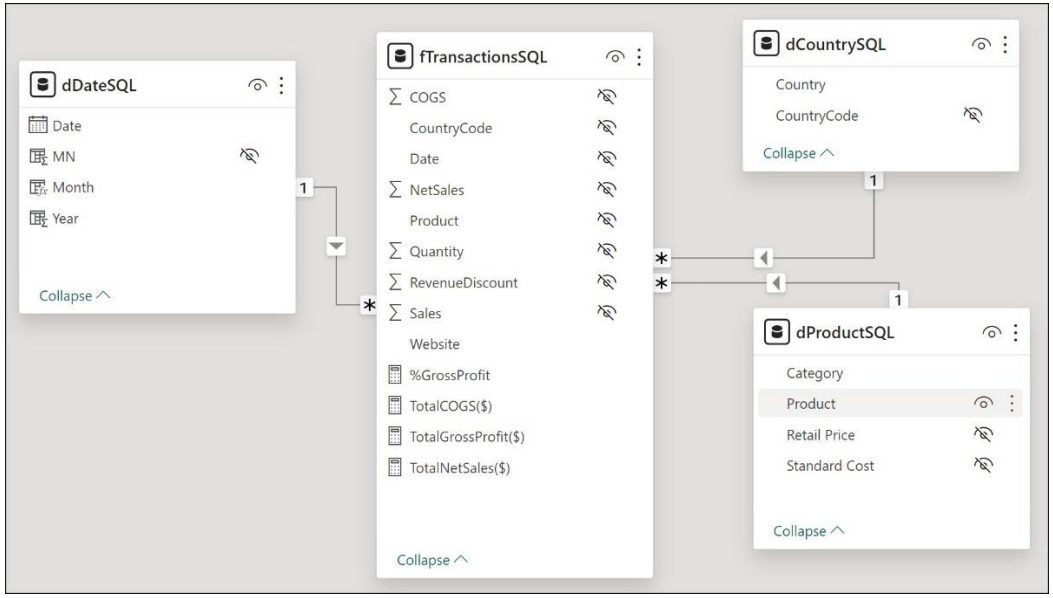
5) **Star schema data model**

A model with a fact table surrounded by dimension tables, relationships, pre-made measures,
and is constructed to be user friendly.

The Data Model in Power Pivot and Power BI are specifically designed to work efficiently with a start schema data model.

*Semantic model in Power BI just means that you upload a model like this, but because it is stored online,

People that are assigned access to the model have: A Single Source of Truth



5) **Steps in build a data model:**

1) Connect to data sources
2) Clean and Transform the data
3) Load to Preferred Location:
    Worksheet
    PivotTable Cache
    Data Model in Power Pivot
    Data Model in Power BI Desktop
    Workspace in Power BI Online
4) Create Relationships (if necessary)
5) Using formulas build metrics and measures
6) Create user friendly reporting area, for example hiding columns

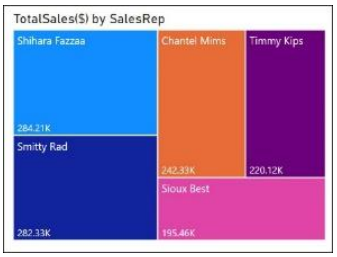
Summary Report

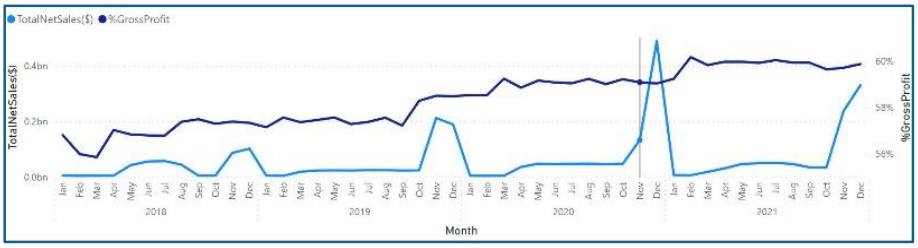
Detailed numbers with labels, which often are metrics that help gauge performance or help make some decision.

Year	TotalSales($)	GrossProfit($)
2024	**665,464.20**	**287,304.26**
Jan	35,200.20	13,833.68
Feb	34,789.20	15,049.16
Mar	97,052.40	41,791.01
Apr	40,824.00	18,415.38
May	38,922.00	16,339.14
Jun	52,399.80	23,603.32
Jul	95,659.80	41,200.72
Aug	37,552.80	14,866.78
Sep	100,695.00	43,665.30
Oct	44,283.60	19,169.41
Nov	56,065.80	25,275.04
Dec	32,019.60	14,095.33
2025	**558,994.80**	**255,659.02**

Visuals

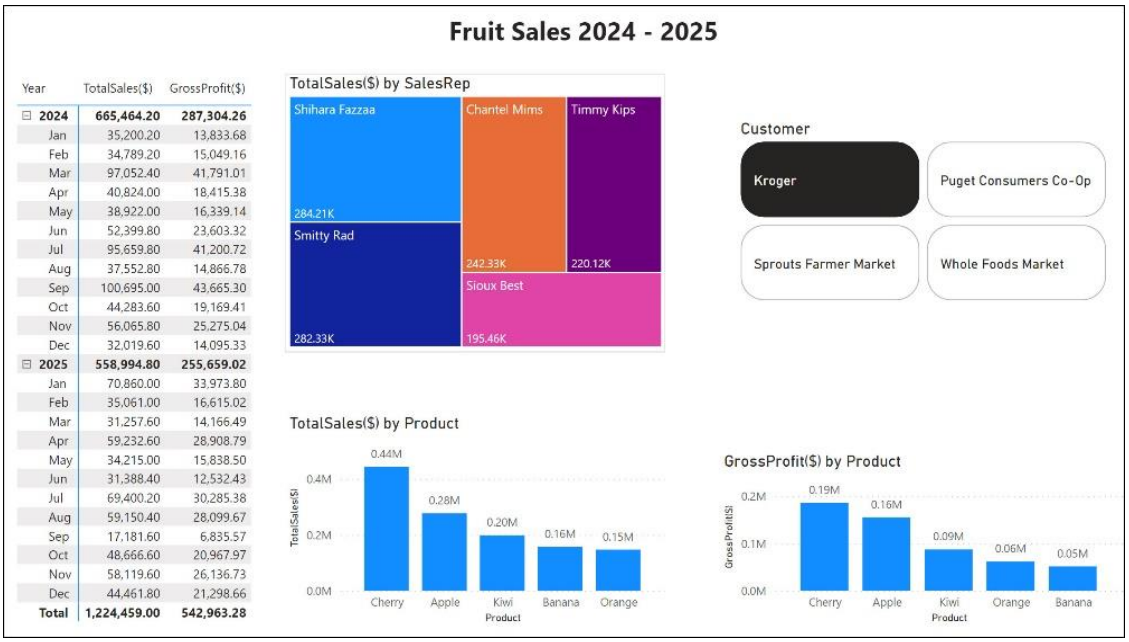
Present quantitative values to get a quick impression, see patterns and trends more quickly than reports or tables.





Dashboards

Reports and visualizations in one location to monitor activity as new data arrives.
Dashboard can be created in Excel, Power BI Desktop, or Power BI Online



Year	TotalSales($)	GrossProfit($)
2024	**665,464.20**	**287,304.26**
Jan	35,200.20	13,833.68
Feb	34,789.20	15,049.16
Mar	97,052.40	41,791.01
Apr	40,824.00	18,415.38
May	38,922.00	16,339.14
Jun	52,399.80	23,603.32
Jul	95,659.80	41,200.72
Aug	37,552.80	14,866.78
Sep	100,695.00	43,665.30
Oct	44,283.60	19,169.41
Nov	56,065.80	25,275.04
Dec	32,019.60	14,095.33
2025	**558,994.80**	**255,659.02**
Jan	70,860.00	33,973.80
Feb	35,061.00	16,615.02
Mar	31,257.60	14,166.49
Apr	59,232.60	28,908.79
May	34,215.00	15,838.50
Jun	31,388.40	12,532.43
Jul	69,400.20	30,285.38
Aug	59,150.40	28,099.67
Sep	17,181.60	6,835.57
Oct	48,666.60	20,967.97
Nov	58,119.60	26,136.73
Dec	44,461.80	21,298.66
Total	**1,224,459.00**	**542,963.28**

Excel	Contains the worksheet, Power Query, M Code, Power Pivot, DAX .
Worksheet Formulas	Worksheet formulas instantly update when the source data changes. No refresh needed.
	You are not confined to structured data such as tables and columns, you easily reference, cells, ranges, columns, or tables.
	You have freedom to incorporate any part of the worksheet and any of the many features.
Legacy worksheet formulas:	Must lock references, copy formulas, edit all cells with formulas and many formulas such as filtering a list are difficult.
Dynamic spilled array formulas	Benefits over Legacy Worksheet Formulas: 1) Offer new array functions such as UNIQUE, SORT and GROUPBY, 2) Usually do not need to lock references, 3) usually do not need to manually copy formula, 4) editing is only done in top cell of array.
Power Query	Case-sensitive function based language, called M Code, to connect to any data source and make any transformation.
	Allow you to work with and transform data structures such as tables, files, columns of files, tables, records or lists.
	Load to worksheet, PivotTable Cache, Data Model, Connection Only, or workspace.
	User interface can write almost all the code for you.
	Memorizes all steps and allows you to go back and change or edit any step.
M Code formulas	Unparalled functional language to work with data to transform and shape into a form that is best suited for the desired analysis output.
PivotTables	For summing, counting and calculating percentages, there is no other calculation tool that is as fast and easy to use as the standard PivotTable.
Power Pivot	Can store millions of rows of data, has great formula advantages with DAX and Relationships, and you can have multiple tables in the reporting area.
DAX formulas	Works with columnar databases to allow calculations across big data. DAX can generate tables of data at any grain internally in the formulas, and thereby reduce the complexities of calculations.
Power BI Desktop	Has all the Data Model advantages, plus it has better visualization capabilities, visuals are interactive, and reports, visuals and dashboards are easy to share.
Power BI Online Service	Advantages of sematic models, reports, visuals, dashboards, Dataflow and workspaces where colleagues can collaborate and connect to a single source of data truth.
Workspace	Area in Power BI Online, where you can assign organizational emails access to the workspace so you can share and collaborate with workspace objects such as: Reports, Workbooks, Dashboards, Semantic Models, Dataflow.
Dataflow	Power Query Online that simultaneously can connect to and transform data, and serve as a single source of data truth.

No	Data Analysis Task	Why use Tools?
1)	**Source data** in worksheet.	
2)	Use **worksheet formulas** to build data model.	Why worksheet formulas for data modeling and reports? Data is already in worksheet, we don't have a lot of data, and calculations we need to make are easy to do with formulas and PivotTables.
3)	**PivotTable** to create year/month sales report.	Why PivotTable? For counting, adding, averaging, % calculations, and year/month totals, PivotTables are easier and faster to use than other tools.
4)	**Spilled Array Formulas** for product sales report.	Why worksheet formulas? Solution instantly updates when source data changes. You can use cells, ranges, columns and tables. PivotTables, M Code and DAX can't do either.
5)	**GROUPBY function** for product sales report.	Why GROUPBY or PIVOTBY? Easier than any other tool, even PivotTable. Solution instantly updates when source data changes. Only works in worksheet.
6)	Txt file, Excel file, **Power Query**, Merge / Join feature to perform lookup. Then Table.Group function to create region report.	Why Power Query? Can work with data structures like tables, files and databases, and can shape data better than other tool. The functions in M Code for dealing with data are unparalleled in the worksheet and DAX. The Merge feature allows you to perform lookup (similar to XLOOKUP and Relationships).
7)	**Changing on-premises path.**	When you connect to a file, the file path is hard coded into M Code formula. If you move the source data file and the destination file to the same new location, the data connection is valid. But if you move the source data file to a new location, you can break the connection. To re-connect, use the Data Source Settings option in There are multiple ways to open this dialog box in Excel and Power BI. If you are in the Power Query Editor: In the Excel Power Query Editor, in the Home tab, Data Sources group, click the Data source settings button. In the Power BI Desktop Power Query Editor, in the Home tab, Data source group, click the Transform data dropdown and then click Data source settings. In the Dataflow Power Query Editor, in the Home tab, Data sources group, click the Manage connections button.

No	Data Analysis Task	Why use Tools?
8)	Duplicate Query, Merge / Join feature to perform lookup, calculated columns in **Power Query, load table of data to PivotTable cache**.	Calculated columns in PQ allow to build full table before loading to PivotTable cache. Loading a table directly to a PivotTable cache prevents the table from being stored both in the worksheet and in the PivotTable cache.
9)	Json file, Excel file, **Power Query, Data Model, DAX** to create reports with SUMX and AVERAGEX. Data Model Relationships to 1) preform lookup, 2) drag fields from dimension tables into reports and visuals to slice and dice.	Data Model and DAX can handle big data and can create tables at different grains internally in formulas more easily than other tools. Average Daily Sales is the example we do to illustrate tables with specified grains in DAX formulas. Relationships help to reduce complexities in formulas, for example the RELATED function for exact match lookup. Relationships also allow you have multiple tables in the reporting area of a PivotTable or Power BI. Filter Context helps formula calculate over big data more efficiently by filtering the large fact table down to just the rows that contain the conditions for the calculation. . Row Context allows formula to see the values in each row of a table, or an iterator function like SUMX. Context Transition is when the row context in a function like AVERAGEX gets converted to filter context to help reduce the number of rows that the fact table has to iterate so that the calculation can be performed efficiently and accurately.
10)	Calculating **average daily units** sold by customer and product.	This is most easily done with DAX formulas because they can determine the grain of a table internally in the formula. Context Transition is when all available row context gets converted to filter context.
11)	Xml file, Excel file, **Power BI Desktop** to build interactive visuals and reports.	Power BI has interactive visuals and reports are easy to share.
12)	Publish report and semantic model to workspace in **Power BI Online**. Connect to the model in the workspace in an Excel file.	Power BI Online allows collaboration in workspaces and can provide semantic models and dataflows as "single source of truth".
13)	**Dataflow** (Power Query Online) to get and transform data and then make it available as a single source of truth to external tools.	Dataflow is like getting two tools for the price of one: 1) get and transform data, 2) Dataflows that serve as a "single source of truth" data source.

No	Data Analysis Task	Why use Tools?
1)	**Source data** in worksheet.	
2)	Use **worksheet formulas** to build data model.	Why worksheet formulas for data modeling and reports? Data is already in worksheet, we don't have a lot of data, and calculations we need to make are easy to do with formulas and PivotTables.
3)	**PivotTable** to create year/month sales report.	Why PivotTable? For counting, adding, averaging, % calculations, and year/month totals, PivotTables are easier and faster to use than other tools.
4)	**Spilled Array Formulas** for product sales report.	Why worksheet formulas? Solution instantly updates when source data changes. You can use cells, ranges, columns and tables. PivotTables, M Code and DAX can't do either.
5)	**GROUPBY function** for product sales report.	Why GROUPBY or PIVOTBY? Easier than any other tool, even PivotTable. Solution instantly updates when source data changes. Only works in worksheet.

Excel Worksheet

=XLOOKUP(LV,MatchC,ReturnC)

PivotTables

Show Values As: % Column Total

dProduct = Excel Table = dimension table

Produc	Produc	CrateP	CrateCost
100CR	Apple	35	15.4
175AP	Orange	29.75	17.255
255YN	Kiwi	22.5	12.6
QA430	Banana	17.5	11.725
BL579	Cherry	43.7	25.346

dPSalesRep= Excel Table = dimension table

SalesR	SalesR	Region
266SB	Sioux Best	West
513CM	Chantel M	West
644SF	Shihara Fa	East
687TK	Timmy Kip	East
699SR	Smitty Rac	MidWest

1) fSales = Excel Table = fact table

Date	Units	Produc	SalesR	Sales	Product	Region
6/19/2025	288	175AP	266SB	8568	Orange	West
10/17/2024	132	100CR	513CM	4620	Apple	West
11/13/2024	132	100CR	266SB	4620	Apple	West
3/31/2025	132	100CR	699SR	4620	Apple	MidWest
10/15/2024	180	100CR	513CM	6300	Apple	West
10/25/2025	156	100CR	699SR	5460	Apple	MidWest
1/15/2024	168	BL579	513CM	7341.6	Cherry	West
8/30/2025	204	255YN	644SF	4590	Kiwi	East
11/24/2024	180	100CR	699SR	6300	Apple	MidWest
3/22/2024	276	QA430	513CM	4830	Banana	West
7/9/2025	108	BL579	644SF	4719.6	Cherry	East
5/26/2025	96	QA430	266SB	1680	Banana	West
8/7/2025	192	100CR	699SR	6720	Apple	MidWest
10/23/2025	156	BL579	644SF	6817.2	Cherry	East
10/12/2024	156	BL579	687TK	6817.2	Cherry	East
9/3/2024	84	255YN	266SB	1890	Kiwi	West
7/1/2025	228	175AP	644SF	6783	Orange	East
9/24/2025	240	175AP	699SR	7140	Orange	MidWest
9/9/2024	72	BL579	266SB			
1/3/2025	216	100CR	644SF			
5/10/2024	84	175AP	266SB			
9/21/2025	96	175AP	513CM			
4/29/2024	216	100CR	644SF			
5/13/2024	180	100CR	513CM			
1/11/2025	300	BL579	699SR			
3/23/2024	48	255YN	266SB	1080	Kiwi	West
10/24/2025	192	175AP	644SF	5712	Orange	East
6/1/2024	216	100CR	699SR	7560	Apple	MidWest
6/13/2025	180	QA430	513CM	3150	Banana	West
8/7/2024	156	BL579	687TK			
2/26/2024	96	BL579	266SB			
7/10/2025	96	QA430	266SB			
8/25/2024	60	255YN	687TK			
10/12/2024	300	QA430	687TK			
5/28/2024	204	255YN	699SR			
1/27/2024	228	100CR	699SR	7980	Apple	MidWest

Worksheet

=SORT(UNIQUE(fSales[Product]))

=SUMIFS(fSales[Sales],fSales[Product],O24#)

=P24#/P29

2) Worksheet formulas for data modeling

3) PivotTable:

Years	Months	Sales ($)	% of Year Total
⊞2024		2,639,338	100.00%
⊟2025	Jan	293,508	11.60%
	Feb	143,626	5.68%
	Mar	182,938	7.23%
	Apr	203,260	8.03%
	May	182,809	7.23%
	Jun	139,240	5.50%
	Jul	254,537	10.06%
	Aug	235,919	9.33%
	Sep	162,038	6.41%
	Oct	238,789	9.44%
	Nov	289,379	11.44%
	Dec	203,795	8.06%
2025 Total		2,529,840	100.00%
Grand Total		5,169,178	

4) Dynamic Spilled Array Formulas:

Product	Total Sales	% Total
Apple	1,260,420	24.4%
Banana	548,940	10.6%
Cherry	1,635,079	31.6%
Kiwi	811,890	15.7%
Orange	912,849	17.7%
Total	5,169,178	100.0%

Worksheet

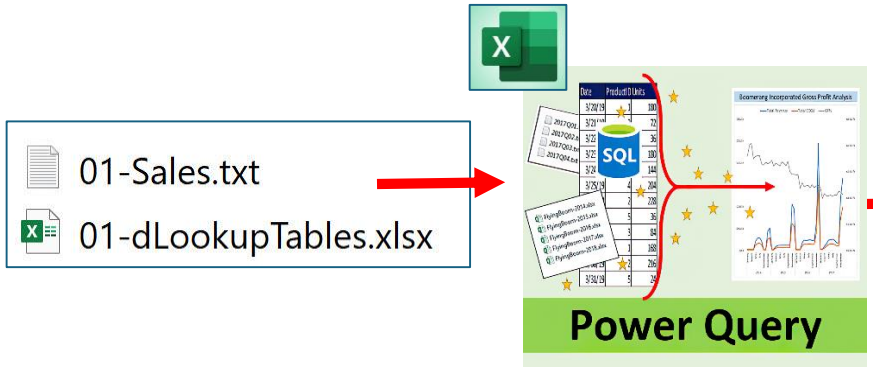
=GROUPBY(fSales[[#All],[Product]], fSales[[#All],[Sales]], HSTACK(SUM,PERCENTOF))

5) GROUPBY Function:

Sales Report:

Product	SUM	PERCENTOF
Apple	1,260,420	24.4%
Banana	548,940	10.6%
Cherry	1,635,079	31.6%
Kiwi	811,890	15.7%
Orange	912,849	17.7%
Total	5,169,178	100.0%

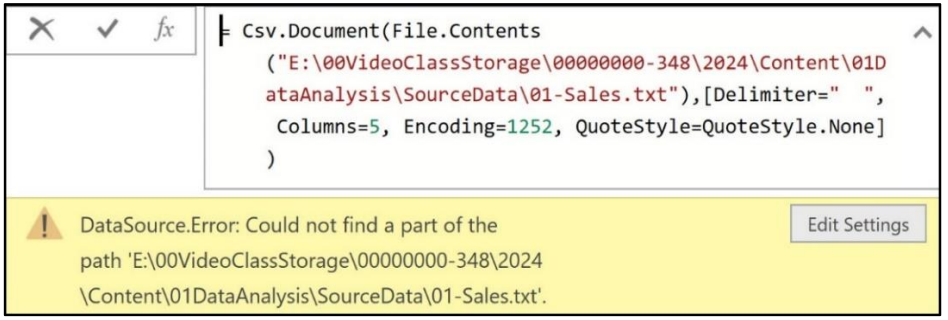
No	Data Analysis Task	Why use Tools?
6)	Txt file, Excel file, **Power Query**, Merge / Join feature to perform lookup. Then Table.Group function to create region report.	Why Power Query? Can work with data structures like tables, files and databases, and can shape data better than other tool. The functions in M Code for dealing with data are unparalleled in the worksheet and DAX. The Merge feature allows you to perform lookup (similar to XLOOKUP and Relationships).
7)	**Changing on-premises path.**	When you connect to a file, the file path is hard coded into M Code formula. If you move the source data file and the destination file to the same new location, the data connection is valid. But if you move the source data file to a new location, you can break the connection. To re-connect, use the Data Source Settings option in There are multiple ways to open this dialog box in Excel and Power BI. If you are in the Power Query Editor: In the Excel Power Query Editor, in the Home tab, Data Sources group, click the Data source settings button. In the Power BI Desktop Power Query Editor, in the Home tab, Data source group, click the Transform data dropdown and then click Data source settings. In the Dataflow Power Query Editor, in the Home tab, Data sources group, click the Manage connections button.



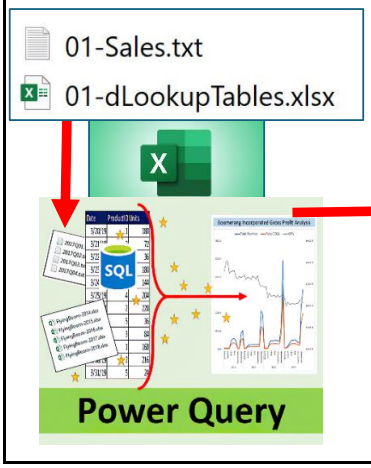
6) Report created by Power Query:

Region	Ave. Units Sold	Standard Deviation
West	156.23	84.03
MidWest	161.67	83.95
East	169.31	82.97

7) On-premises file path error:



No	Data Analysis Task	Why use Tools?
8)	Duplicate Query, Merge / Join feature to perform lookup, calculated columns in **Power Query, load table of data to PivotTable cache**.	Calculated columns in PQ allow to build full table before loading to PivotTable cache. Loading a table directly to a PivotTable cache prevents the table from being stored both in the worksheet and in the PivotTable cache.



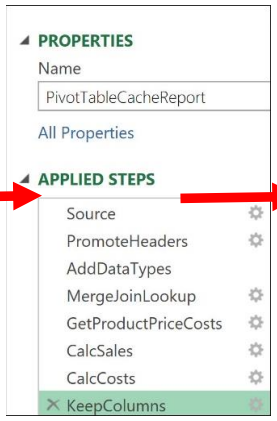
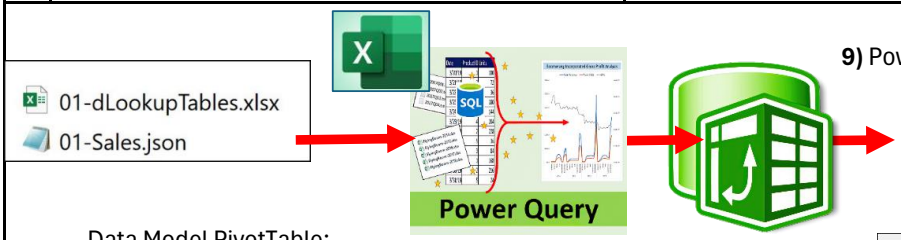


Table loaded to behind-the scenes **PivotTable Cache**

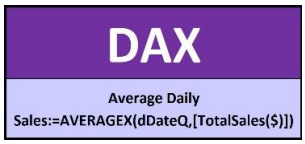
8) Load Power Query data to PivotTable:

Product	Sales ($)	% of Total Sales	Costs($)	% of Cost Totals
Apple	1,260,420	24.38%	554,585	19.43%
Banana	548,940	10.62%	367,790	12.88%
Cherry	1,635,079	31.63%	948,346	33.22%
Kiwi	811,890	15.71%	454,658	15.93%
Orange	912,849	17.66%	529,452	18.55%
Grand Total	**5,169,178**	**100.00%**	**2,854,831**	**100.00%**

No	Data Analysis Task	Why use Tools?
9)	Json file, Excel file, **Power Query, Data Model, DAX** to create reports with SUMX and AVERAGEX. Data Model Relationships to 1) preform lookup, 2) drag fields from dimension tables into reports and visuals to slice and dice.	Data Model and DAX can handle big data and can create tables at different grains internally in formulas more easily than other tools. Average Daily Sales is the example we do to illustrate tables with specified grains in DAX formulas. Relationships help to reduce complexities in formulas, for example the RELATED function for exact match lookup. Relationships also allow you have multiple tables in the reporting area of a PivotTable or Power BI. Filter Context helps formula calculate over big data more efficiently by filtering the large fact table down to just the rows that contain the conditions for the calculation. . Row Context allows formula to see the values in each row of a table, or an iterator function like SUMX. Context Transition is when the row context in a function like AVERAGEX gets converted to filter context to help reduce the number of rows that the fact table has to iterate so that the calculation can be performed efficiently and accurately.
10)	Calculating **average daily units** sold by customer and product.	This is most easily done with DAX formulas because they can determine the grain of a table internally in the formula. Context Transition is when all available row context gets converted to filter context.



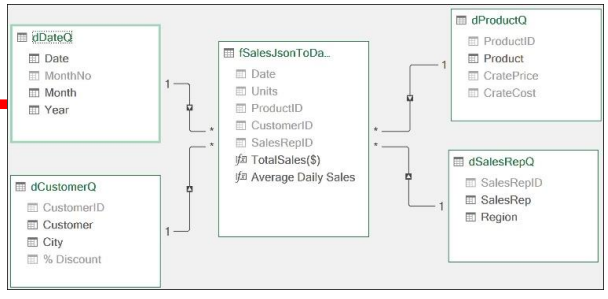
9) Power Pivot's Data Model - DAX Measures & Relationships:



Data Model PivotTable:

10) Average is made at day grain:

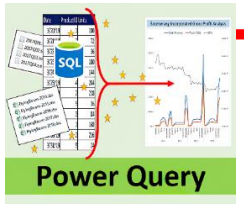
Product	TotalSales($)	Average Daily Sales
Apple	1,260,420.00	6,463.69
Banana	548,940.00	3,248.17
Cherry	1,635,079.20	8,054.58
Kiwi	811,890.00	4,460.93
Orange	912,849.00	5,246.26
Grand Total	**5,169,178.20**	**9,484.73**



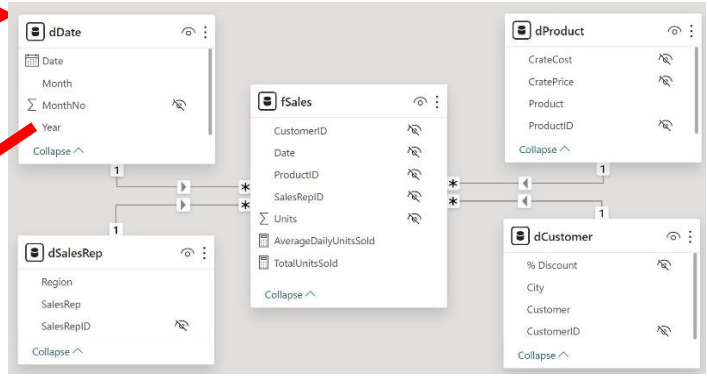
No	Data Analysis Task	Why use Tools?
11)	Xml file, Excel file, **Power BI Desktop** to build interactive visuals and reports.	Power BI has interactive visuals and reports are easy to share.

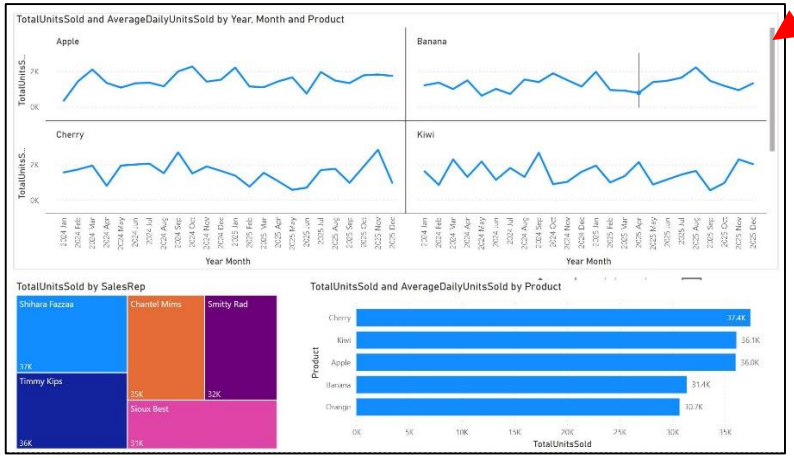
Sales.xml
01-dLookupTables.xlsx

Power BI **Desktop**



11) Build Power BI Data Model and interactive visuals:

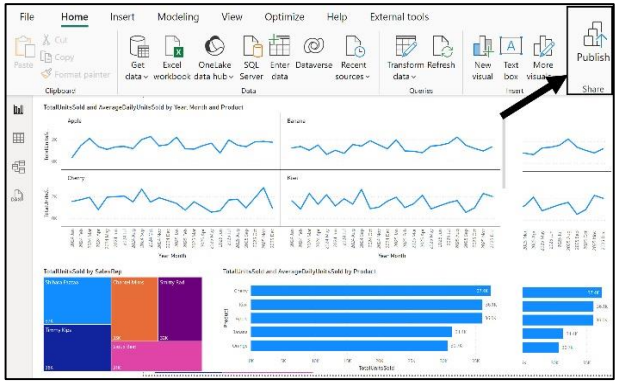




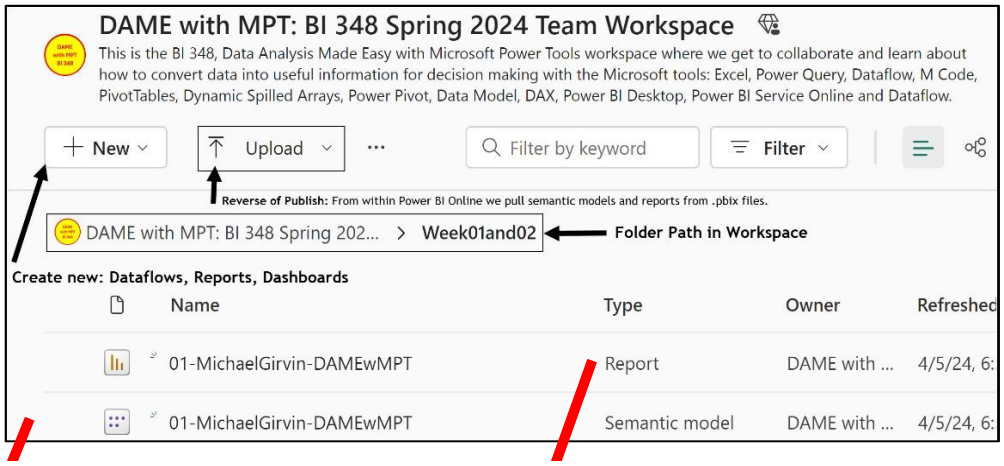
No	Data Analysis Task	Why use Tools?
12)	Publish report and semantic model to workspace in **Power BI Online**. Connect to the model in the workspace in an Excel file.	Power BI Online allows collaboration in workspaces and can provide semantic models and dataflows as "single source of truth".

Power BI **Desktop**

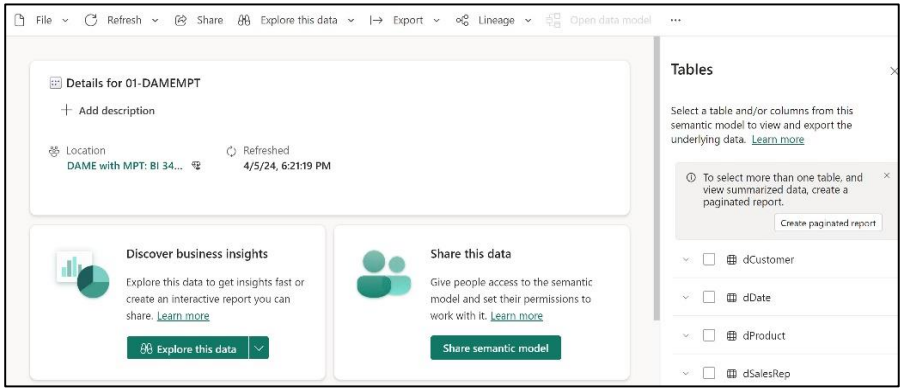
12) Publish to Workspace:



Power BI **Online Service**

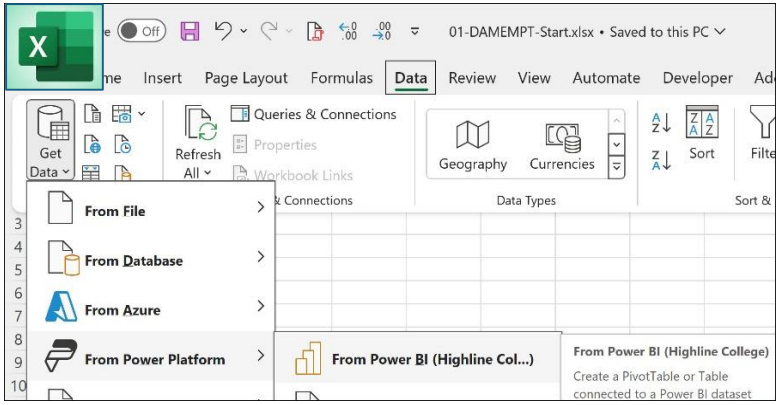


Data Model gets published as a Semantic Model (Single Source of Truth):

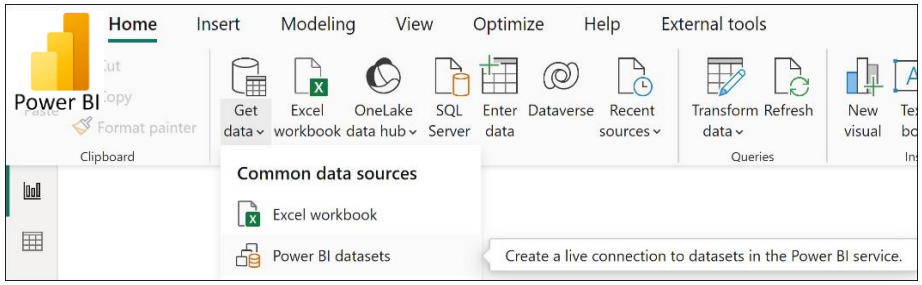


Report gets published as Report that anyone in workspace an consume:

Connect to semantic model from Excel:

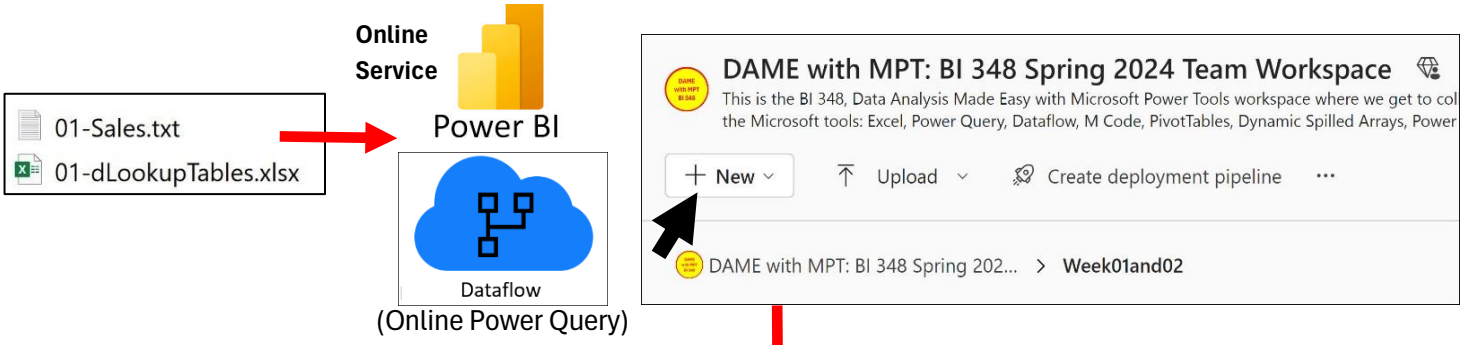


Connect to semantic model from Power BI Desktop:

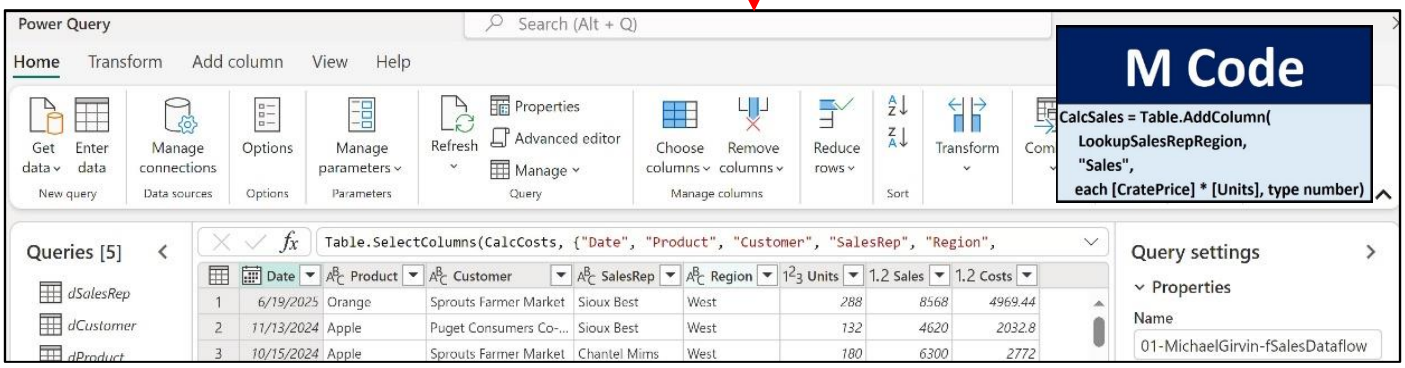


No	Data Analysis Task	Why use Tools?
13)	**Dataflow** (Power Query Online) to get and transform data and then make it available as a single source of truth to external tools.	Dataflow is like getting two tools for the price of one: 1) get and transform data, 2) Dataflows that serve as a "single source of truth" data source.

13) Dataflow in Power BI Online Workspace:

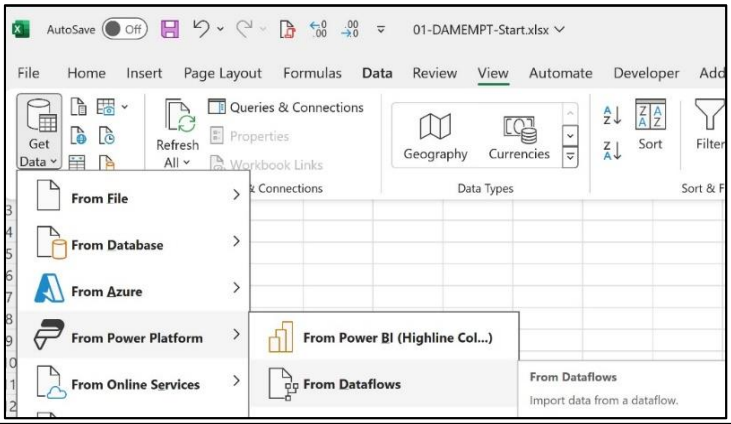


1) Get, transform data, save to workspace:

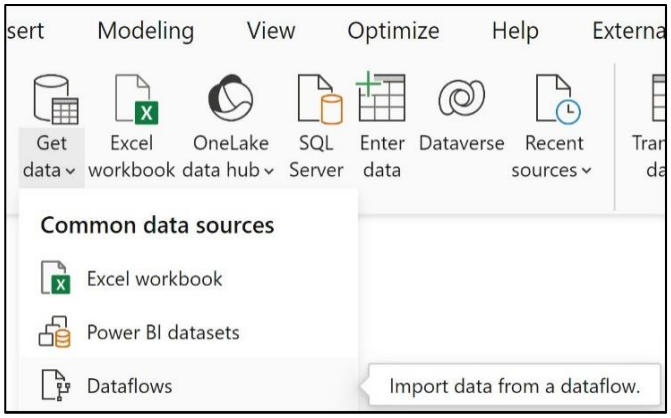


2) "Single source of truth". Workspace Dataflows are data source for external tools like Excel and Power BI Desktop:

Excel connects to Dataflows:



Power BI Desktop connects to Dataflows:



When we use Dataflow to upload files, the files are stored in OneDrive. Here are the OneDrive status icons that you need to be aware of:

Icon	Icon Name	Description
	Blue cloud	file doesn't download to your device until you open it. You can't open online-only files when your device isn't connected to the Internet.
	Shared with People	Indicates the file or folder has been shared with other people.
	Online, Locally Downloaded	When you open an online-only file (blue cloud), it downloads to your device and becomes a locally available file. If you need more space, you can change the file back to online only: right-click the file and select "Free up space." With Storage Sense turned on, these files will become online-only files after the time period you've selected.
	Always keep on this device	These files download to your device and take up space, but they're always there for you even when you're offline. OneDrive is just keeping a copy online as backup.
	Padlock	File or folder has settings which prevent it from syncing.
	New File	File or folder is new. You'll see this when using OneDrive.com online.
	Red X	File or folder cannot be synced. You'll see this in File Explorer or on the OneDrive notification area icons.
	Learn More	Click the OneDrive icon in the notification area to learn more about the problem.
	Gray OneDrive icon	Means you're not signed in, or OneDrive setup hasn't completed.
	Pasue	The paused symbol over the OneDrive icon means your files are not currently syncing. To resume syncing, select the OneDrive icon in the notification area, select More and then Resume syncing.
	Sync pending	The circular arrows over the OneDrive notification icons signify that sync is in progress. This includes when you are uploading files, or OneDrive is syncing new files from the cloud to your PC.
	Account blocked	Account is blocked.
	Warning	Your account needs attention. Select the icon to see the warning message displayed in the activity center.
	Two Cloud Icon	Signed into both work or school and a personal account. The blue one is for your work or school account, the white one is for your personal account.

Data and table terms:

Variable	A value that can change, like a product name or an amount of a sale. Variables can be quantitative (number), categorical (text), Boolean (T/F), or other.
Data	Values collected for one or more variables and kept together for reference or analysis. Data stored in its smallest form. **Data:** 22 | 3/29/2024 | Product **Not data:** 22,3/29/2024, Product Data is not information. Information is created from data.
Data Type	Declared type of data for a column such as: number, text or logical. Safeguards for the column of data and helps to assure data consistency and accuracy for reports and visuals.
Field	Column that is used to collect data for a variable. Column should have declared data type and must have field name at top of column. In the world of databases, columns are called fields. In other arenas, such as Microsoft power tools, fields are often called columns.
Field Name	Name at top of field that accurately describes the data. Field names are used in reports and visuals to indicate which variables to summarize Synonyms = column name = header name.
Record	A row in table that contains related data for each field for a given observation, such as a sales transaction, scientific observation or employee data.
Table	A collection of one or more columns, with field names in first row, records of data are in subsequent rows, and data types for each column. Data must be contained in a table to easily be analyzed. A great amount of data is not stored in tables, very often, the job of the analysts is to transform the unstructured data into a table structure.
Fact Table	A table that contains the facts to summarize or measure, like sales, units, amount of time, sports statistics, and other facts. The facts are the measurements of activities, like amount of sales, number of points scored in a game or length of time at a web site. Fact tables are sometimes very large and can have thousands, millions or billions of rows of data.
Dimension (Lookup) Table	A table that contains a field with unique list of entities, called a primary key column, and subsequent columns with attributes for the entity. An entity like Product ID would have attributes like product name, price, cots and product weight. Dimension tables are also called lookup tables.
Primary Key	The entity field in a dimension table that contains a unique list of items and is used to assure that there are no duplicate records in the dimension table.
Foreign Key	When a primary key is used in a fact table it is called a foreign key.
One-To-Many Relationship	When the primary key from a dimension table is connected to a foreign key in a fact table, the primary is called the one-side and the foreign key is called the many-side. A one-to-many relationship helps to make lookup formulas easy and allows dimension table attributes to filter reports and visuals.
Note:	The terms "fact" and "dimension" go back to 1960s when the General Mills company and Dartmouth University used them to name their tables of data. Then in the 1970s the data research companies AC Nielsen & IRI company used the terms. However, the terms were popularized by Ralph Kimball in the 1980s with his extensive writing about data warehousing and business intelligence.
Grain or Granularity	The level of detail stored in a table, or the size of the number. A table of invoice total sales amounts has a larger grain and less detail than a table of line items sales amounts. A table of line items sales amounts has a smaller grain and more detail than a table of invoice total sales amounts. The grain of the 2023 total sales amount is bigger and has less detail than the grain of the May, 2024 sales amount.

	The grain of the May, 2024 sales amount is smaller and has more detail than grain of the 2023 total sales amount. This concept is important because if you have two fact tables at different grains, you often have to create a method to convert the two fact tables into one.
Database	Location where most data in the world is stored.
Relational Database	A database that follows strict rules for storing related tables of data with no redundancy.
Text Files	A common vehicle to transfer tables of data from one system to another system using delimiters such as comma and tab.
Source data	The original location of the data, like in a text file, an Excel file of a database.
Data destination	The location where the data is loaded, such as in an Excel or a Power BI Desktop file or an online source like a Power BI workspace.
On-premises file path	A hard coded source data file path in the data destination, such as an Excel or Power BI Desktop file. On-premises file paths can cause errors when the data destination file is moved and the connection to the source data is lost.
Online source data	Online source data can solve the problem of On-premises file and folder paths. Web sites, SQL Server databases and Power BI Online are examples of online sources that stay connected to the source data when the data destination file is moved.
Delimiter	Is a character that separated bits of data, such as a comma, tab and other characters.
Structure Or Schema	The rules or structure for tables, data files and databases.
Data Analysis Business Intelligence	Converts data into useful information to gain insight and make decisions. Information can be in the form of: reports, visuals, dashboards, and other forms. Data analysis allows you to make data-driven decisions, which tend to be more accurate & help to achieve goals more consistently Synonyms: Data Analytics, Analytics, Business Intelligence, Data Science. The definition is the same as data analysis, but the process is performed within the context of business data and business decision making.
Data analysis process	1. Determine what questions need answers and what decisions need to be made. Everything else in the process is dictated by these questions and decisions. 2. Where is the data? How much data? What is the structure of the data? 3. Which MS tool to use? (Almost always starts with Power Query). 4. Clean, transform and shape the data into a table or model that is best suited to answer questions and make decisions, 5. Build final model with measures, metrics, relationships, and other features. 6. Create useful information: reports, visuals and dashboards. 7. Refresh when new data arrives. 8. Change and update model as necessary.
Query	A question that we ask of the raw data, like import Csv file, append tables, or group line item sales to calculate invoice total sales.
Data modeling Cleaning data Transforming data	Involves all the steps necessary to get and transform the data from the original data sources into a data structure that allows you to create the useful information required. Data modeling can be done with many tools such as: Power Query, worksheet formulas, DAX formulas, Relationships and many other Excel and Power BI features. Involves fixing individual bites of data, such as extracting text from a larger text string, converting text dates to proper dates or rounding numbers. Involves structural changes to source data or tables of data, such as converting Json files to proper tables, merging tables to add columns, or converting two fact tables into one table.
M Code (Data Mashup)	Case-sensitive function language in Power Query that allows you to work with data and data structures to clean, transform and shaped the data into the required structure.
A data model	Is the final data structure that allows you to create the useful information required is called a data model and becomes the intermediate step between the source data and the reports and visuals.
Star schema data model	A model with a fact table surrounded by dimension tables, relationships between the tables (usually one-to-many), pre-made measures, and is constructed to be user friendly. The Data Model in Power Pivot and Power BI are specifically designed to work efficiently with a start schema data model.

Semantic model in Power BI	It is the same as a Data Model in Power Pivot and Power BI Desktop, and is usually a Star schema data model. The difference is that the model is stored online, contains data security features and serves as a single source of online data truth.
The Data Model	Is the location in Power Pivot and Power BI Desktop where the data is stored in a columnar database and the DAX formulas, relationships and other model features are added.
Columnar database	In the Data Model, behind the scenes, there is a RAM memory Columnar Database that compresses and stores the data, and which allows the DAX formulas to work efficiently with big data.
DAX (Data Analysis eXpressions)	Functional language used in the Data Model to work efficiently with big data and can internally create tables at different grains to reduce the complexity of many formulas.
Measures	Metrics that help gauge performance or help make some decision.
One-to-many Relationship	Primary key of dimension table is connected to the foreign key in the fact table to simplify lookup formulas and allow dimension table fields in the report area to filter reports, visuals and databases.
Load data	Involves deciding where to load the data, such as loading to the worksheet, a PivotTable Cache, the Data Model, or a Power BI workspace.
Summary Report	Reports often contain detailed numbers with labels, and the numbers are almost always metrics that help gauge performance or help make some decision. Reports almost always contains conditional calculations and therefore it is very helpful if you are fluent with logical tests.
Visuals	Present quantitative values in a visual way to get a quick impression and see patterns and trends more quickly than reports or tables . Examples of visuals: column, bar, line, scatter, map, waterfall charts, Pictures, Conditional Formatting, and more.
Dashboards	Contain summary reports, visualizations, and other elements in one location so that we can monitor the activity as new data arrives. Just like a dashboard in a car, a dashboard should present information that is required for making good decisions
Excel Worksheet Formulas	Contains the worksheet, Power Query, M Code, Power Pivot, DAX . Worksheet formulas instantly update when the source data changes. No refresh needed. You are not confined to structured data such as tables and columns, you easily reference, cells, ranges, columns, or tables. You have freedom to incorporate any part of the worksheet and any of the many features.
Legacy worksheet formulas:	Must lock references, copy formulas, edit all cells with formulas and many formulas such as filtering a list are difficult.
Dynamic spilled array formulas	Do NOT need to lock references or copy formulas, edit is done in top cell of array only, and formulas for filtering are simple.
Power Query	Case-sensitive function based language to connect to any data source and make any transformation. Allow you to work with and transform data structures such as tables, files, columns of files, tables, records or lists. Load to worksheet, PivotTable Cache, Data Model, or workspace. User interface can write almost all the code for you. Memorizes all steps and allows you to go back and change or edit any step.
PivotTables	For summing, counting and calculating percentages, there is no other calculation tool that is as fast and easy to use as the standard PivotTable.
Power Pivot	Can store millions of rows of data, has great formula advantages with DAX and Relationships, and you can have multiple tables in the reporting area.
Power BI Desktop	Has all the Data Model advantages, plus it has interactive visualizations.
Power BI Online Service	Advantages of sematic models, reports, visuals, dashboards, Dataflow and workspaces where colleagues can collaborate and connect to a single source of data truth.
Workspace	Area in Power BI Online, where you can assign organizational emails access to the workspace so you can share and collaborate with workspace objects such as: Reports, Workbooks, Dashboards, Semantic Models, Dataflow.
Dataflow	Power Query Online that simultaneously can connect to and transform data, and serve as a single source of data truth.