Data Analysis & Business Intelligence Made Easy with Excel Power Tools
Excel Data Analysis Basics = E-DAB
Notes for Video:
E-DAB-02: Data, Proper Data Sets, Excel Tables, Sort, Filter, Logical Tests

Objectives of Video:

1. What is Excel?

Ribbon Tabs provide access to Commands that we can click with our Mouse. The Data Ribbon Tab is selected to see groups of data-related commands
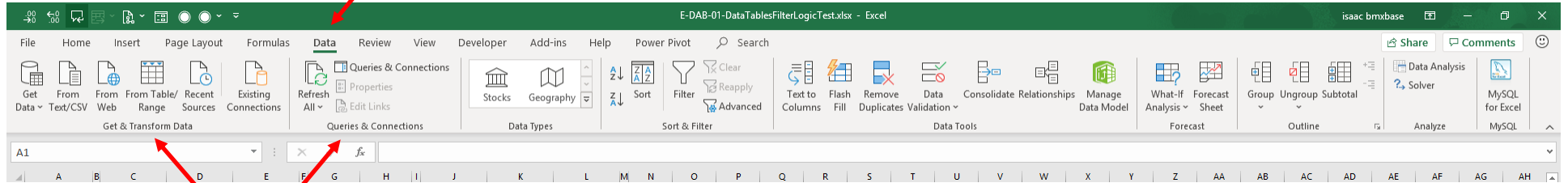
Groups such as "Get & Transform Data" and "Queries & Connections" provide Commands that we can click with our Mouse. The "Get & Transform Data" group contains Commands that open the Power Query editor. The "Get & Transform Data" and "Queries & Connections" groups are often referred to as the query tool "Power Query".

2. What is Data, Raw Data?

Objective #2: What is Data, Raw Data?

Raw Data = **Data** = Data stored in its smallest form that allows Excel Data Analysis Features & Power Tools to work.

Not Raw Data: | Raw Data stored in smallest form, stored in a Proper Data Set. | Not Raw Data. | Data in Proper Data Set

Addresses
313 173rd Blvd, Kent, WA 98121
316 66th Blvd, Kent, WA 98124
4358 23rd St, Kent, WA 98122
965 151st St, Kent, WA 98116
7900 173rd Lane, Kent, WA 98126
4047 15th Ave, Kent, WA 98122
4907 13th Ave, Kent, WA 98123
3789 4th Blvd, Seattle, WA 98115
2977 66th Lane, Seattle, WA 98117
3392 23rd St, Seattle, WA 98113

Address	City	State	Zip
313 173rd Blvd	Kent	WA	98121
316 66th Blvd	Kent	WA	98124
4358 23rd St	Kent	WA	98122
965 151st St	Kent	WA	98116
7900 173rd Lane	Kent	WA	98126
4047 15th Ave	Kent	WA	98122
4907 13th Ave	Kent	WA	98123
3789 4th Blvd	Seattle	WA	98115
2977 66th Lane	Seattle	WA	98117
3392 23rd St	Seattle	WA	98113

Date.Product.Sales
2/15/2019-Quad-287.7
2/15/2019-Aspen-89.85
2/15/2019-Aspen-239.6
2/15/2019-Sunspot-65.85
2/15/2019-Aspen-329.45
2/15/2019-Quad-383.6
2/15/2019-Sunspot-131.7
2/15/2019-Sunspot-285.35
2/15/2019-Quad-239.75
2/15/2019-Quad-239.75

Date	Product	Sales
2/15/19	Quad	287.7
2/15/19	Aspen	89.85
2/15/19	Aspen	239.6
2/15/19	Sunspot	65.85
2/15/19	Aspen	329.45
2/15/19	Quad	383.6
2/15/19	Sunspot	131.7
2/15/19	Sunspot	285.35
2/15/19	Quad	239.75
2/15/19	Quad	239.75

Data Analysis Features & Tools won't work on Data stored this way.

Data Analysis Features & Tools work on Data stored this way. Easier to analyze data when it is stored in its smallest form.

Hard to add sales by Product

Easy to add sales by Product

Example:
Hard to sort this data by Zip Code.

Example:
Easy to sort this data by Zip Code.

3. Understanding the Difference Between Data & Information.

Objective #3: Difference Between Data & Information?

Data = Unorganized Raw Data
Data alone doesn't yield insight
Data alone is not very useful

Information = Organized & Presented Data
Helps people make decisions
Helps see patterns and gain insight

Define Data Analysis = Convert Data Into Useful Information

Raw **Data**:

Date	Product	Sales ($)
2/15/2019	Quad	287.7
2/15/2019	Aspen	89.85
2/15/2019	Aspen	239.6
2/15/2019	Sunspot	65.85
2/15/2019	Aspen	329.45
2/15/2019	Quad	383.6
2/15/2019	Sunspot	131.7
2/15/2019	Sunspot	285.35
2/15/2019	Quad	239.75
2/15/2019	Quad	239.75

Convert Into ==>>

Useful **Information**:

Goal: Which Product Had Lowest Sales?

Answer = Information Created from **Raw Data**

Product	Sum of Sales
Aspen	$658.90
Quad	$1,150.80
Sunspot	$482.90
Grand Total	$2,292.60

Raw **Data**:

Date	Product	Sales ($)
Mar	Quad	560
Apr	Quad	781
May	Quad	612
Jun	Quad	714
Jul	Quad	932
Aug	Quad	717
Sep	Quad	718
Oct	Quad	1203
Nov	Quad	1445
Dec	Quad	1334

Convert Into ==>>

Useful **Information**:

Goal: Trend in Quad Sales?

Answer = Information Created from **Raw Data**

Quad Sales ($) 2019

560 781 612 714 932 717 718 1203 1445 1334

Mar Apr May Jun Jul Aug Sep Oct Nov Dec
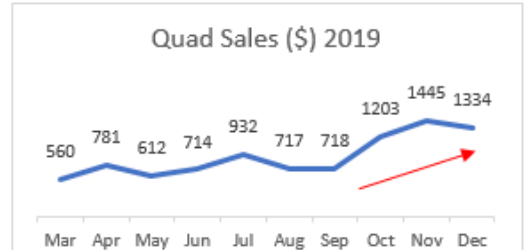
4. Define Proper Data Set.

Objective #4: Define Proper Data Set.

Proper Data Set = Table = Table of Data = Table with 1) Field Names in first row and 2) Records containing Data in subsequent rows.

In Excel you must have empty cells or Column/Row headers around the Proper Data Set in order for the Excel & Power Tools to work.

Fields = Column Headers =

Other terms you may see: Field Names, Columns, Variables, Attributes

Fields define what type of data goes into the column

Data Type =

What type of data is in Column? Date, Number, Text and so on?

Traditionally, in an Excel Spreadsheet we do not have Data Types.

Power Tools (Power Query, Power Pivot and Power BI Desktop) we will have to define Data Types if we want our analysis to work correctly.

Records = Rows = One Data point for each Field in a row.

Examples of records:

1) Sales Transactions, where each row represents a product sold with data like date and sales amount

2) Unique Product Information, where each row has the product code, price, cost and other product attributes.

Delimiter = In Text Files (Special Files that hold Proper Data Sets) a Delimiter is a designated character, such as a Tab or Comma, that separates each Field and each Small Piece of Raw Data.

Naming = Name all tables, fields and other items: 1) smartly and 2) consistently

Example 1: Proper Data Set in Excel. Records = Sales Transaction

Date	ProductID	Units Sold	Sales ($)
2/15/19	1043	6	287.7
2/15/19	1069	3	89.85
2/15/19	1069	8	239.6
2/15/19	2005	3	65.85
2/15/19	1069	11	329.45
2/15/19	1043	8	383.6
2/15/19	2005	6	131.7
2/15/19	2005	13	285.35
2/15/19	1043	5	239.75
2/15/19	1043	5	239.75

Fields, 1st row

Records with Data in each row

Example 2: Proper Data Set in Excel. Records = Unique Product

ProductID	Product	Flight Range (M)	Retail Price
1043	Quad	20	47.95
1069	Aspen	25	29.95
2005	Sunspot	15	21.95

When a Table of Data is in an Excel Worksheet, there must be empty cells all the way around data set, or Excel Column Row Headers

5. <u>Why Proper Data Sets are Mandatory.</u>

- There are least four reasons that Proper Data Sets are mandatory for Data Analysis:
    1. All Data Analysis Tools that we use require that the Data is stored in a Proper Data Set. If the Data is not stored in a Proper Data Set, we can NOT use the tools to analyze our Data. If this is the case, we will have to clean and transform the Data into a Proper Data Set.
    2. Almost all data storage systems in the world use Proper Data Sets to store Raw Data.
    3. It makes sense that we use this form to store and analyze data, because it requires that we keep similar data in one column.
    4. It further makes sense because when we analyze our data, we ask questions of or query one column at a time, and so having similar data in clearly named columns in a Proper Data Set makes the analysis much easier than if we stored the data in a different form.

6. <u>Tables are not Charts</u>

<div style="border: 2px solid red; padding: 10px;">

Objective #6: Tables are not Charts.

Table = Field Names & Records with Data.   **Chart** = Visual Portrayal of Number Data.

Date	Product	Sales ($)
Mar	Quad	560
Apr	Quad	781
May	Quad	612
Jun	Quad	714
Jul	Quad	932
Aug	Quad	717
Sep	Quad	718
Oct	Quad	1203
Nov	Quad	1445
Dec	Quad	1334

Quad Sales ($) 2019

560 781 612 714 932 717 718 1203 1445 1334

Mar Apr May Jun Jul Aug Sep Oct Nov Dec
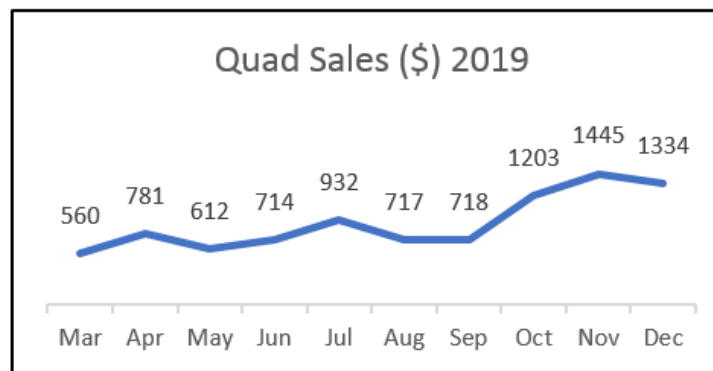
We are displaying how a number changes over the category "Month"

We are storing Data in a Proper Data Set (Table)

This may seem obvious, but people confuse these two items regularly in academia, statistics, and other fields

</div>

7. Use Excel Tables For Dynamic Data.

	A	B	C	D	E	F	G	H	I	J
1										
2		**Objective #7: Use Excel Tables For Dynamic Data.**								
3										
4		We can covert Proper Data Sets that are located in an Excel Worksheet to an Excel Table using the								
5		Excel Table feature with these steps:								
6		1) Click in 1 cell in the Proper Data Set.								
7		2) Use the keyboard: Ctrl + T (or Ribbon Tab: Insert, Tables, Table) to Convert to an Excel Table.								
8		3) ALWAYS name your Table (Keyboard: Alt, J, T, A or Ribbon Tab: TableTools-Design, Properties, Table Name)								
9		Three Reasons we use Excel Tables rather than a Proper Data set in an Excel Worksheet:								
10		1) They provide Dynamic Ranges. This means that when we add or remove records from the table, anything pointing								
11		to the Excel Table (Formulas, PivotTables, Power Query, Power Pivot, Charts) can be updated when source data changes.								
12		2) Data from an Excel Worksheet MUST be in an Excel Table BEFORE it can be imported into Power Query.								
13		3) Data from an Excel Worksheet MUST be in an Excel Table BEFORE it can be imported into Power Pivot.								
14										
15		**Proper Data Sets in Worksheet:**								
16										
17		Date	ProductID	Units Sold	Sales ($)		ProductID	Product	Flight Range (M)	Retail Price
18		2/15/19	1043	6	287.7		1043	Quad	20	47.95
19		2/15/19	1069	3	89.85		1069	Aspen	25	29.95
20		2/15/19	1069	8	239.6		2005	Sunspot	15	21.95
21		2/15/19	2005	3	65.85					
22		2/15/19	1069	11	329.45					
23		2/15/19	1043	8	383.6					
24		2/15/19	2005	6	131.7					
25		2/15/19	2005	13	285.35					
26		2/15/19	1043	5	239.75					
27		2/15/19	1043	5	239.75					
28										
29		**Proper Data Sets converted to Excel Tables and give a Table Name.**								
30										
31		**fTransactions Excel Table:**					dProduct Excel Table:			
32										
33		Date	ProductID	Units Sold	Sales ($)		ProductID	Product	Flight Range (M)	Retail Price
34		2/15/19	1043	6	287.7		1043	Quad	20	47.95
35		2/15/19	1069	3	89.85		1069	Aspen	25	29.95
36		2/15/19	1069	8	239.6		2005	Sunspot	15	21.95
37		2/15/19	2005	3	65.85					
38		2/15/19	1069	11	329.45					
39		2/15/19	1043	8	383.6					
40		2/15/19	2005	6	131.7					
41		2/15/19	2005	13	285.35					
42		2/15/19	1043	5	239.75					
43		2/15/19	1043	5	239.75					

8. Data Types in Proper Data Sets in Various Tools

Throughout our study of Data Analysis, we will use various methods and tools to store our Proper Data Sets. In addition, each of these methods and tools have a unique way of handling different Data Types when they exist in a single column. Below are 8 examples of the different situations we will encounter.

	A	B	C	D	E	F	G
1	**Examples of Proper Data Sets:**						
2							
3			**1) Excel Worksheet.**				
4			We do NOT have to explicitly declare the Data Type for each Column.				
5			Inconsistent Data Types can exist in same column.				
6							
7			Date	ProductID	Units Sold	Sales ($)	
8			2/15/2019	1043	6	287.7	
9			2/15/2019	1069	3	89.85	
10			2/15/2019	1069	8	239.6	
11			2/15/2019	2005	3	65.85	
12	Number Date ==>>		2/15/2019	1069	11	329.45	
13	Text Date ==>>		15/2/2019	1043	8	383.6	
14			2/15/2019	2005	6	131.7	
15			2/15/2019	2005	13	285.35	
16			2/15/2019	1043	5	239.75	
17			2/15/2019	1043	5	239.75	
18							

	A	B	C	D	E	F	G
1	**Examples of Proper Data Sets:**						
18							
19			**2) Excel Table (using Excel Table Feature)**				
20			We do NOT have to explicitly declare the Data Type for each Column.				
21			Inconsistent Data Types can exist in same column.				
22							
23			Date	ProductID	Units Sold	Sales ($)	
24			2/15/2019	1043	6	287.7	
25			2/15/2019	1069	3	89.85	
26			2/15/2019	1069	8	239.6	
27			2/15/2019	2005	3	65.85	
28	Number Date ==>>		2/15/2019	1069	11	329.45	
29	Text Date ==>>		15/2/2019	1043	8	383.6	
30			2/15/2019	2005	6	131.7	
31			2/15/2019	2005	13	285.35	
32			2/15/2019	1043	5	239.75	
33			2/15/2019	1043	5	239.75	
34							

Examples of Proper Data Sets:

3) Power Query editor window
We MUST explicitly declare the Data Type for each Column in the Power Query editor window.
Inconsistent Data Types can NOT exist in same column. If it does, we get an Error.

Number Date ==>>
Error ==>>

	Date	123 ProductID	123 Units Sold	1.2 Sales ($)
1	2/15/2019	1043	6	287.7
2	2/15/2019	1069	3	89.85
3	2/15/2019	1069	8	239.6
4	2/15/2019	2005	3	65.85
5	2/15/2019	1069	11	329.45
6	Error	1043	8	383.6
7	2/15/2019	2005	6	131.7
8	2/15/2019	2005	13	285.35
9	2/15/2019	1043	5	239.75
10	2/15/2019	1043	5	239.75

fTransactionsPQTable - Power Query Editor

Examples of Proper Data Sets:

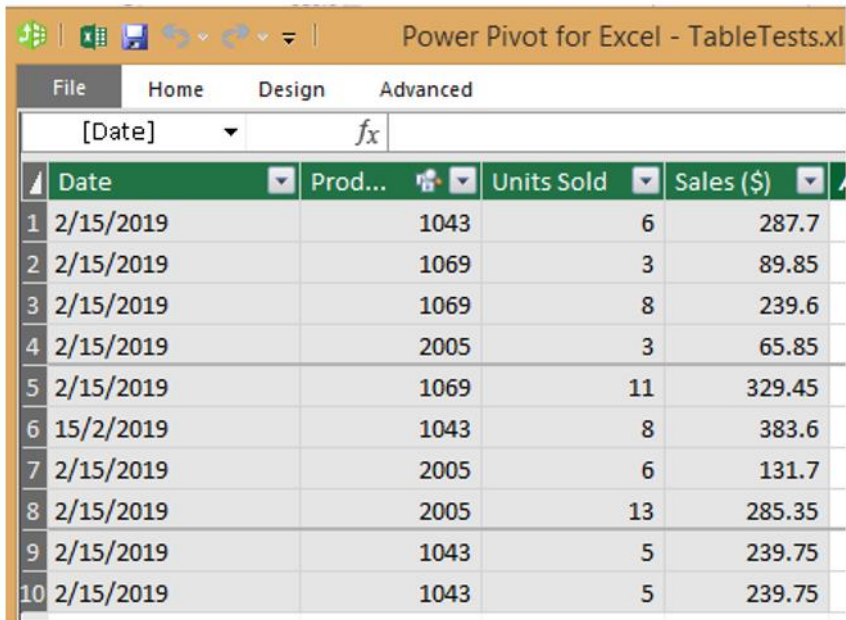### 4) Power Query Output when it is Loaded to Excel Worksheet as an Excel Table
The Data Type was explicitly declared in the Power Query editor window.
Inconsistent Data Types can NOT exist in same column. If it does we get an Empty Cell.

Number Date ==>>
Empty Cell ==>>

Date	ProductID	Units Sold	Sales ($)
2/15/2019	1043	6	287.7
2/15/2019	1069	3	89.85
2/15/2019	1069	8	239.6
2/15/2019	2005	3	65.85
2/15/2019	1069	11	329.45
	1043	8	383.6
2/15/2019	2005	6	131.7
2/15/2019	2005	13	285.35
2/15/2019	1043	5	239.75
2/15/2019	1043	5	239.75

5) Excel Power Pivot Window
We MUST explicitly declare the Data Type for each Column in Power Pivot.
Inconsistent Data Types can NOT exist in same column.
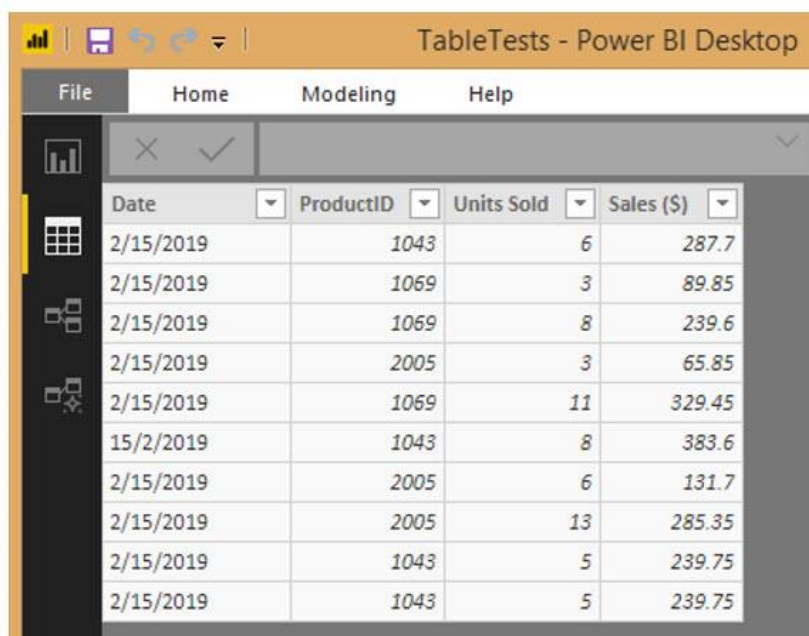
All Data is
One Data Type ==>>

6) Excel Power BI Desktop
We MUST explicitly declare the Data Type for each Column in Power BI Dektop.
Inconsistent Data Types can NOT exist in same column.

All Data is
One Data Type ==>>

Examples of Proper Data Sets:

7) Text Files

Text Files are used to transfer Proper Data Sets from one system to another.

All Data is Text and must be converted to the correct Data Type.

when we import these text files.

All Text Data Type =>

```
Date → ProductID → Units ·Sold→ Sales ·($)¶
2/15/2019 → 1043 → 6    →   287.7¶
2/15/2019 → 1069 → 3    →   89.85¶
2/15/2019 → 1069 → 8    →   239.6¶
2/15/2019 → 2005 → 3    →   65.85¶
2/15/2019 → 1069 → 11   →   329.45¶
15/2/2019 → 1043 → 8    →   383.6¶
2/15/2019 → 2005 → 6    →   131.7¶
2/15/2019 → 2005 → 13   →   285.35¶
2/15/2019 → 1043 → 5    →   239.75¶
2/15/2019 → 1043 → 5    →   239.75¶
```

Examples of Proper Data Sets:

8) Fact Tables and Dimension Tables

Fact Tables have the Numbers we need to summarize (like Sales or Units)

Date	ProductID	Units Sold	Sales ($)
2/15/19	1043	6	287.7
2/15/19	1069	3	89.85
2/15/19	1069	8	239.6
2/15/19	2005	3	65.85
2/15/19	1069	11	329.45
2/15/19	1043	8	383.6
2/15/19	2005	6	131.7
2/15/19	2005	13	285.35
2/15/19	1043	5	239.75
2/15/19	1043	5	239.75

Dimension Tables have unique lists of attributes used as conditions to make the Fact Table calculations

ProductID	Product	Flight Range (M)	Retail Price
1043	Quad	20	47.95
1069	Aspen	25	29.95
2005	Sunspot	15	21.95

9. <u>How to Sort Data.</u>

Objective #9: How to Sort Data.

What does Sorting do?

 Organizes a Proper Data Set (Table) or a single column in alphabetical or numeric order.

To Sort, use the drop-down arrow

 A to Z (Small to Big, Ascending).

 Z to A (Big to Small, Descending).

* When you sort a Proper Data Set, records remain intact.

* To sort more than one column, sort the "Major Sort Column" last.

* Sorting works the same in an Excel Table, Power Query, Power Pivot and Power BI Desktop

Product	SalesRep	Sales ($)
Quad	Gigi	239.75
Aspen	Gigi	239.60
Sunspot	Gigi	131.70
Sunspot	Gigi	65.85
Quad	Sioux	239.75
Aspen	Sioux	89.85
Quad	Tyrone	383.60
Aspen	Tyrone	329.45
Quad	Tyrone	287.70
Sunspot	Tyrone	285.35

This Table is sorted first by Sales Z to A, then by SalesRep A to Z

This Table is sorted first by the Sales ($) Column, Z to A, then by SalesRep Column, the "Major Sort Column", A to Z.

For the SalesRep Tyrone, we can see that the Sales ($) Column is sorted, biggest to smallest, for his records.

10. Filtering & Extracting Data.

Objective #10: Filtering & Extracting Data.

What does Filter do?

Filtering allows you to selects rows based on conditions or criteria.

In Excel, the Filter feature allows us to hide records that don't match criteria, then copy & paste matched records to a different location.

In Power Query, when we filter, it removes records that do not match the condition or criteria.

To Filter, use the drop-down arrow:

Filter dropdown Arrows always shows a Unique List of Items from the Field/Column

Filter allows you to filter in these ways:

Check Box from Unique List

Built-in features for the Data Types: Date, Text, Numbers

You can use an AND Logical Test, OR Logical Test, or other types of Logical Test

** Perfect Feature to Extract Records that match a set of criteria.

Filter Example #1: Select "Alma" from Unique List of Items to show only Alma records. Copy and Paste to new sheet.

Filter Example #2: AND Logical Test: Show Records for Sales Rep "Alma" AND Autos "Chevy".

Filter Example #3: OR Logical Test: Show Records for Sales Rep "Alma" OR "Rina".

Filter Example #4: Number Filter to show top 5 sales records.

Date	Sales Rep	Autos	Sales
4/30/17	Alma	Chevy	$88,016
9/27/16	Alma	Chevy	$69,598
1/30/17	Alma	Chevy	$107,034

Using the drop-down arrows to select the item "Alma" from the Sales Rep Column and then "Chevy" from the Autos Column, we are showing only the records for Sales Rep "Alma" AND " Auto "Chevy". This is an example of an AND Logical Test because in order for the record to show, we needed the record to contain "Alma" and "Chevy".

In Excel, we can see that many rows that did not match our AND Logical Test are hidden

11. Understand & Use AND Logical Tests and OR Logical Tests

1) AND Logical Test (using AND Criteria):
    1. You can have two or more criteria for an AND Logical Test.
    2. If we select the check the boxes for "Alma" on the Sales Rep Field and "Chevy" on the Auto Field:
        - For each record we are asking two questions:
            1. "Is the Sales Rep *Alma*?"
            AND
            2. "Is the Auto sold **Chevy**?"
        - For each Record we can get these possible answers:
            1. TRUE, FALSE
            2. FALSE, TRUE
            3. FALSE, FALSE
            4. TRUE, TRUE.
    3. For an AND Logical Test you must get "**All Are TRUE**", in order for the record to be included in the filtered data set.
    4. AND Logical Tests are used in all aspects of Data Analysis and in all Data Analysis Tools. Although we saw how to run an AND Logical Test for filtering, throughout the class we will see then in other situations like adding with an AND Logical test and in other tools such as PivotTables, SUMIFS Spreadsheet Functions, Power Query and more.
2) OR Logical Test (using OR Criteria):
    1. You can have two or more criteria for an OR Logical Test.
    2. If we select the check the boxes for "Alma" and "Rina" in the Sales Rep Field:
        - For each record we are asking two questions:
            1. "Is the Sales Rep *Alma*?"
            OR
            2. "Is the Sales Rep *Rina*?"
        - For each Record we can get these possible answers:
            1. TRUE, FALSE
            2. FALSE, TRUE
            3. FALSE, FALSE.
    3. For an OR Logical Test you must get "**At Least 1 TRUE**", in order for the record to be included in the filtered data set.
    4. For Filtering, when we are asking the OR Criteria Question, we are often asking the question of only ONE Column.
    5. OR Logical Tests are used in all aspects of Data Analysis and in all Data Analysis Tools. Although we saw how to run an OR Logical Test for filtering, throughout the class we will see then in other situations like adding with an OR Logical test and in other tools such as PivotTables, Power Query, Power Pivot and more.
3) Other Logical Tests we will use this the class:
    1. BETWEEN Logical Test = an AND Logical Test with an Upper and Lower Limit.
    2. NOT Logical Test = Items are included in the final result when the item is NOT equal to the condition.
    3. CONTAINS Logical Test = Items are included when the designated field contains the condition, like with the CONTAINS Condition "Cola", from the list of items "Coca Cola", "7-Up", "RC Cola", only the items "Coca Cola" and "RC Cola" would match the CONTAINS Condition "Cola".

4) Below is an example of using an AND Logical Test for Filtering and for Adding Sales in a PivotTable Report.

	A	B	C	D	E	F	G	H	I	J	K
1											
2		**Objective #11: Logical Tests**									
3											

Objective #11: Logical Tests

Date	Sales Rep	Autos	Sales
4/30/2017	Alma	Chevy	$88,016
9/27/2016	Alma	Chevy	$69,598
1/30/2017	Alma	Chevy	$107,034

Using Filtering, we are showing only the records for Sales Rep "Alma" AND " Auto "Chevy". This is an example of an AND Logical Test because in order for the record to show, we needed the record to contain "Alma" and "Chevy".

Sum of Sales ($)	Autos				
Sales Rep	**Chevy**	**Ford**	**Honda**	**Toyota**	**Grand Total**
Alma	264,648	33,105		125,940	423,693
Jaeyoung			171,395		171,395
Rina	33,415			47,352	80,767
Tyrone	134,274	97,432	71,880		303,586
Grand Total	**432,337**	**130,537**	**243,275**	**173,292**	**979,441**

Using a PivotTable, we are adding the sales only for the records where the Sales Rep is "Alma" AND the Auto is "Chevy". This is an example of an AND Logical Test because in order for the sales number to be added, we needed the record to contain "Alma" and "Chevy".