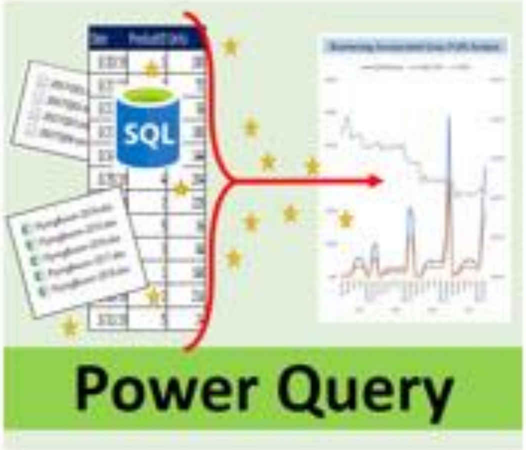
Data Analysis & Business Intelligence Made Easy with Excel Power Tools
Excel Data Analysis Basics = E-DAB
Notes for Video:
E-DAB 08: Import SQL Big Data into Power Pivot & Build DAX Formulas

Outcomes for Video:

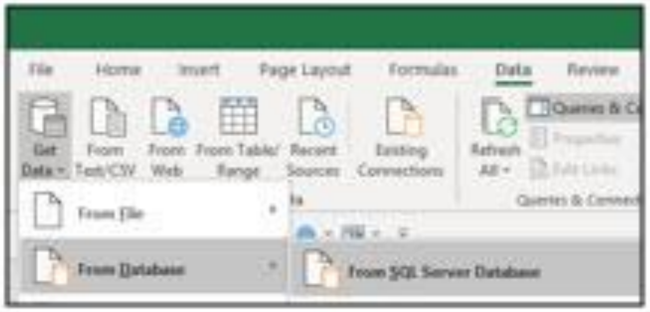
1. Over View of Video Project:

1) Source Data:

7.7 Million Rows of Transational Sales Record



2) Import, Clean and Transform:



Data Model PivotTable

Data Model Pivot Chart

3) Load to:

Excel Power Pivot's **Data Model**:

1) Behind the scens RAM Memory **Columnar Database**

Allows compression of Big Data

(7 millions rows compressed to 25 Megabytes)

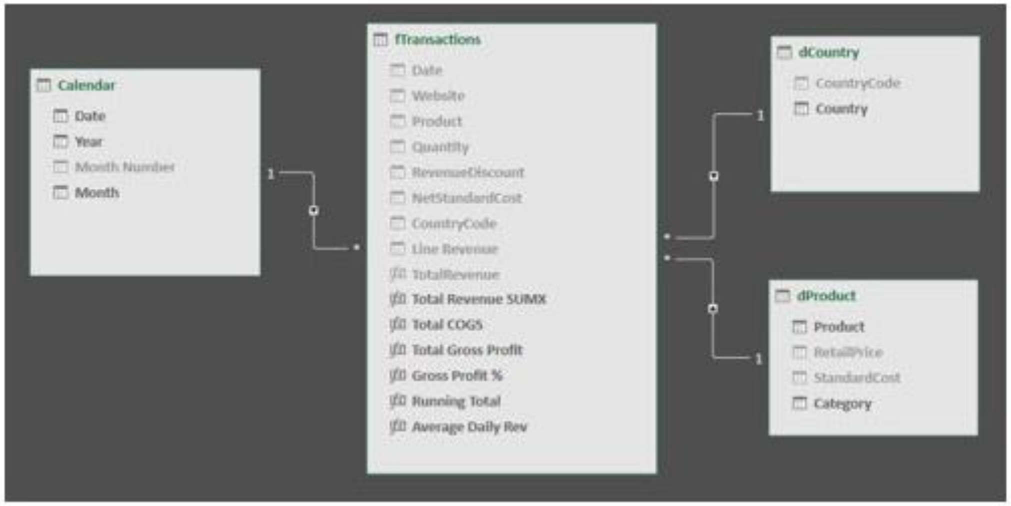
4) Build Data Model:

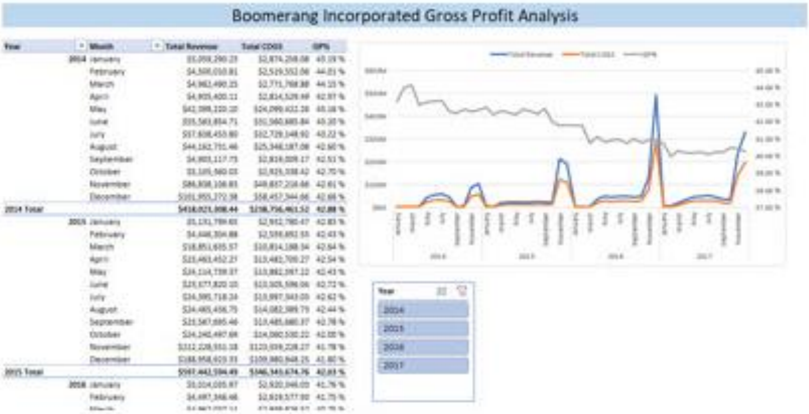
2) Relationships

3) Hide Columns not Used for Reporting

4) Explicit DAX Formulas

5) Build Useful Information (Reports & Visiuals)




Boomerang Incorporated Gross Profit Analysis

Year	Month	Total Revenue	Total COGS	GP%
2014	January	$5,059,290.23	$2,874,258.08	43.19 %
	February	$4,508,010.81	$2,519,552.06	44.11 %
	March	$4,962,490.25	$2,771,768.88	44.15 %
	April	$4,905,400.11	$2,814,529.49	42.87 %
	May	$42,399,220.10	$24,099,422.38	43.16 %
	June	$35,563,854.71	$20,160,680.84	43.31 %
	July	$37,608,455.80	$21,729,148.92	42.22 %
	August	$44,162,755.46	$25,346,187.08	42.60 %
	September	$4,903,117.75	$2,818,009.17	42.53 %
	October	$5,105,560.03	$2,925,338.42	42.70 %
	November	$86,308,106.83	$49,537,218.66	42.61 %
	December	$101,955,272.38	$58,457,344.66	42.66 %
2014 Total		$418,021,908.44	$238,756,461.52	42.88 %
2015	January	$5,131,789.63	$2,932,780.47	42.85 %
	February	$4,446,304.88	$2,559,892.55	42.43 %
	March	$18,851,605.57	$10,814,188.34	42.64 %
	April	$23,463,452.27	$13,482,700.27	42.54 %
	May	$24,114,739.37	$13,882,397.12	42.43 %
	June	$23,077,820.10	$13,305,596.06	42.72 %
	July	$24,395,718.24	$13,997,343.03	42.62 %
	August	$24,465,456.75	$14,082,389.79	42.44 %
	September	$23,547,695.46	$13,485,880.37	42.78 %
	October	$24,240,497.68	$14,080,530.22	42.08 %
	November	$212,228,551.18	$123,559,228.27	41.78 %
	December	$188,958,619.31	$109,980,948.25	41.80 %
2015 Total		$597,442,594.49	$346,345,674.76	42.03 %
2016	January	$5,014,035.87	$2,920,346.03	41.76 %
	February	$4,497,346.46	$2,619,577.92	41.75 %
	March	[illegible]	[illegible]	[illegible]

Year

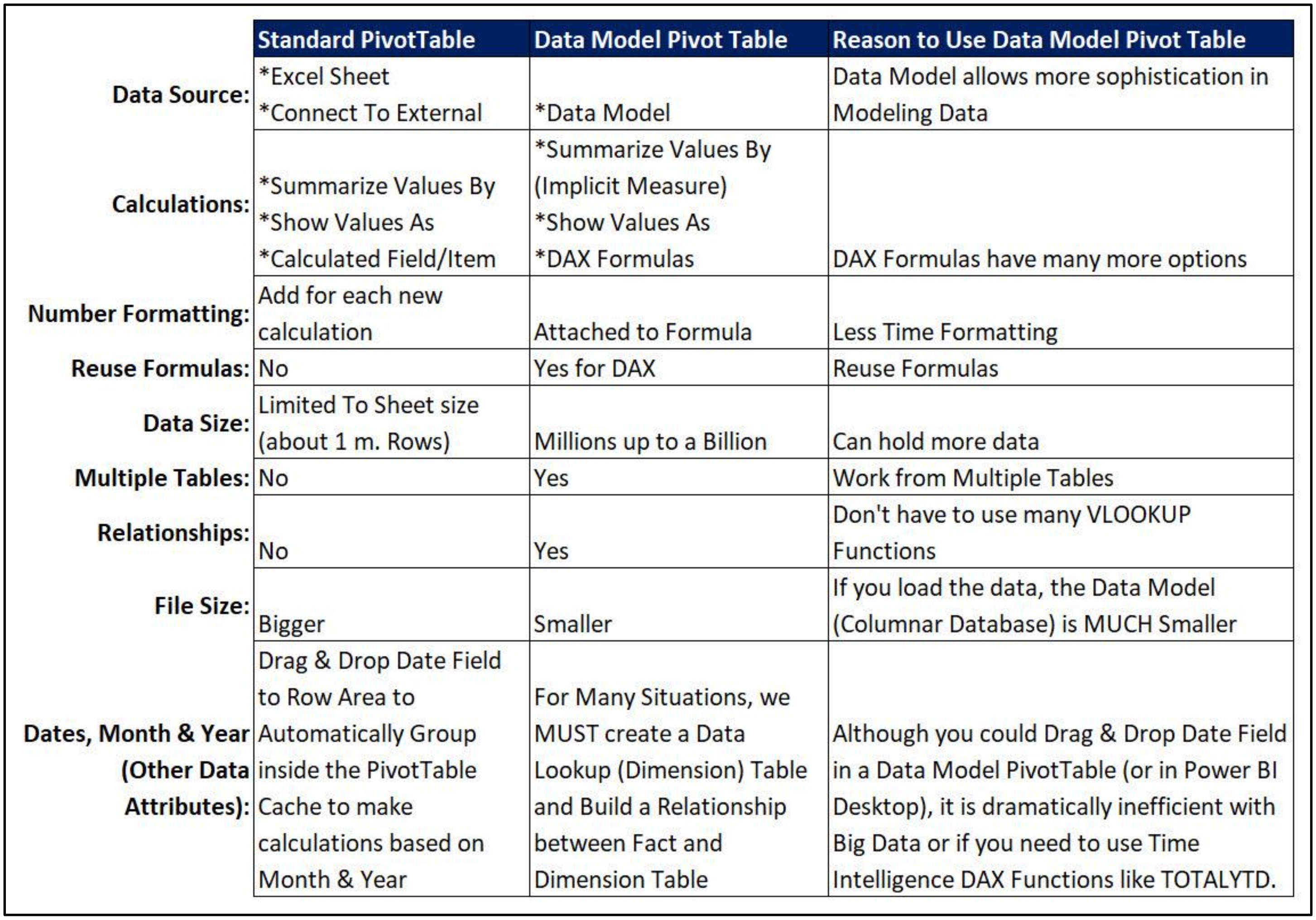
2014

2015

2016

2017

2. Differences Between Standard PivotTable & Data Model PivotTable



	Standard PivotTable	Data Model Pivot Table	Reason to Use Data Model Pivot Table
Data Source:	*Excel Sheet *Connect To External	*Data Model	Data Model allows more sophistication in Modeling Data
Calculations:	*Summarize Values By *Show Values As *Calculated Field/Item	*Summarize Values By (Implicit Measure) *Show Values As *DAX Formulas	DAX Formulas have many more options
Number Formatting:	Add for each new calculation	Attached to Formula	Less Time Formatting
Reuse Formulas:	No	Yes for DAX	Reuse Formulas
Data Size:	Limited To Sheet size (about 1 m. Rows)	Millions up to a Billion	Can hold more data
Multiple Tables:	No	Yes	Work from Multiple Tables
Relationships:	No	Yes	Don't have to use many VLOOKUP Functions
File Size:	Bigger	Smaller	If you load the data, the Data Model (Columnar Database) is MUCH Smaller
Dates, Month & Year (Other Data Attributes):	Drag & Drop Date Field to Row Area to Automatically Group inside the PivotTable Cache to make calculations based on Month & Year	For Many Situations, we MUST create a Data Lookup (Dimension) Table and Build a Relationship between Fact and Dimension Table	Although you could Drag & Drop Date Field in a Data Model PivotTable (or in Power BI Desktop), it is dramatically inefficient with Big Data or if you need to use Time Intelligence DAX Functions like TOTALYTD.

3. Choice between: Standard PivotTable & Data Model PivotTable :

Standard PivotTable:

1. Have One Flat Table
2. Don’t have Big Data
3. Standard Calculation in PT sufficient
4. Must manually add Number Format for each new Calculation
5. Can NOT re-use a Formula
6. For simple PivotTable Reports on a small data set, Standard PivotTables are great.

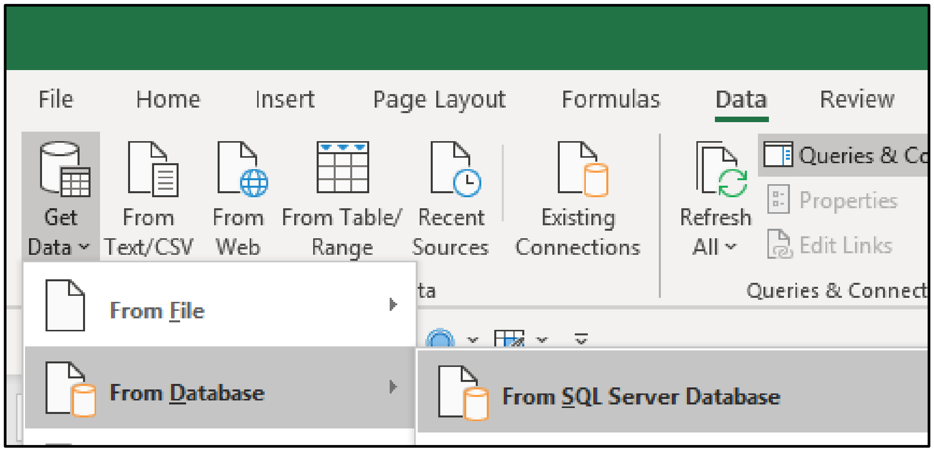
Data Model PivotTable:

1. Have Multiple Tables
2. Have Big Data
3. More Varied Calculations with DAX
4. Number Formatting can be added to formula
5. DAX Measures (Formulas) are created once, and can be re-used many times
6. For complex projects or Big Data, Data Model PivotTables are great.

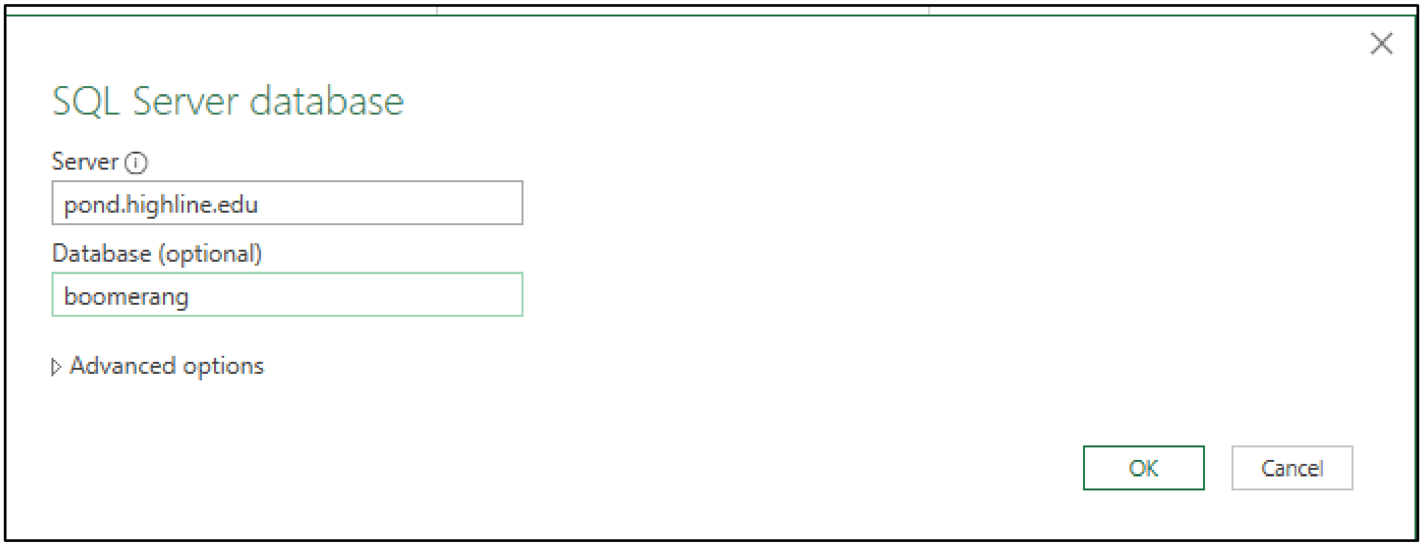
I switch from Standard PivotTable over to Data Model PivotTables when these occur:

1. Have more than one table
2. Have more than about 50,000 rows of data
3. Want DAX Formulas because:
   a. DAX can make more varied calculations
   b. Can have Number Formatting attached to formula
   c. Can use formula over and over.

4. **SQL Database :**
   - An SQL Database is a common database where tables of raw data are stored.
   - An SQL database is a relational database that uses SQL Code (Structured Query Language) to query or communicate with the database.
   - Power Query can easily connect to an SQL Database and import tables of data.
5. **Power Query to Import from SQL Database** :
   - Use the Get Data dropdown to select the SQL Database as seen here:



   - Enter Credentials, as seen here:



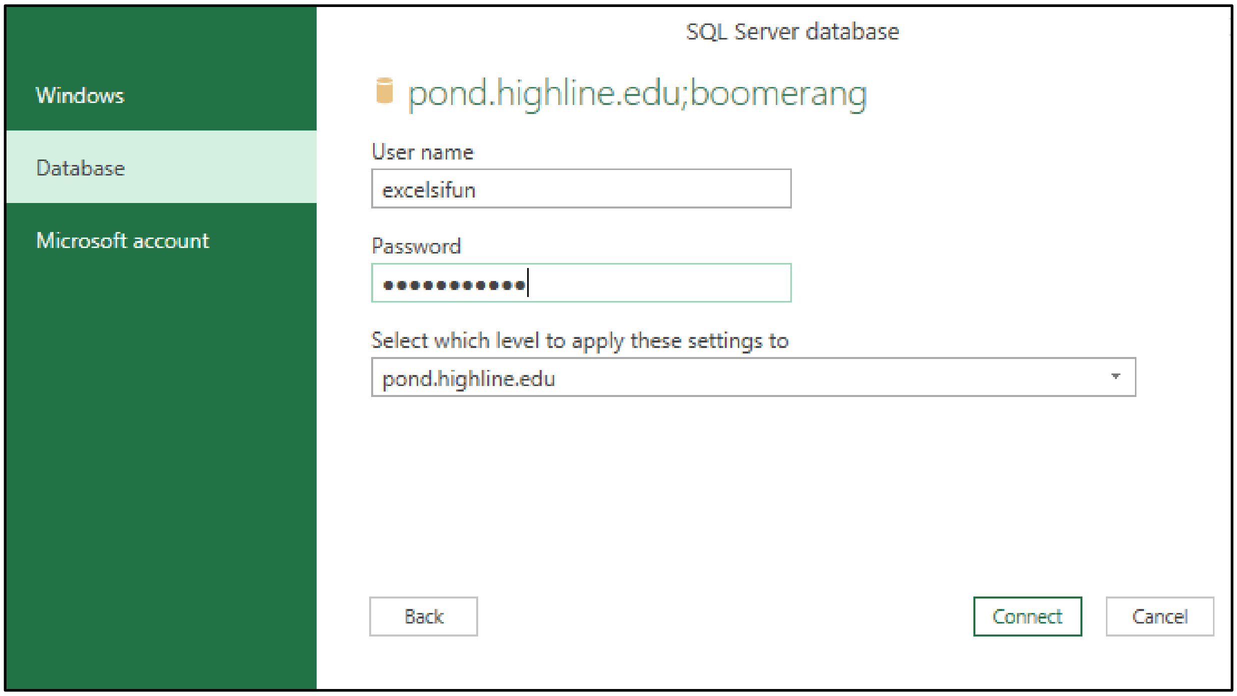
SQL Server database

Server: pond.highline.edu

Database (optional): boomerang

   - In next window select the Database option and then enter the user and password as seen here:

Server:	pond.highline.edu
Database:	boomerang
User:	excelisfun
Password:	ExcelIsFun!



SQL Server database

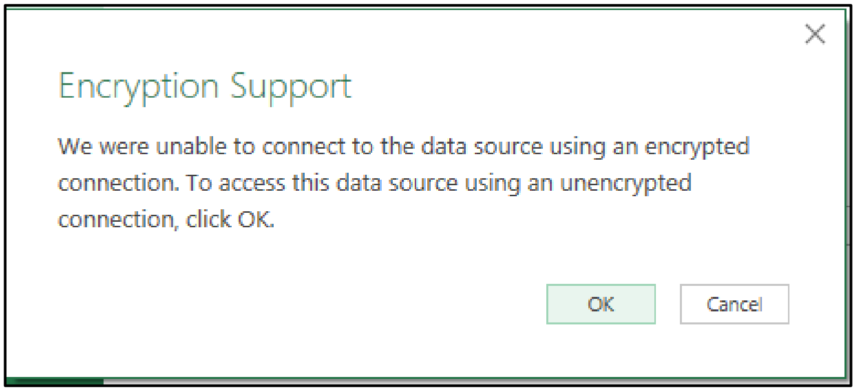
pond.highline.edu;boomerang

User name: excelsifun

Password: ●●●●●●●●●●●

Select which level to apply these settings to: pond.highline.edu

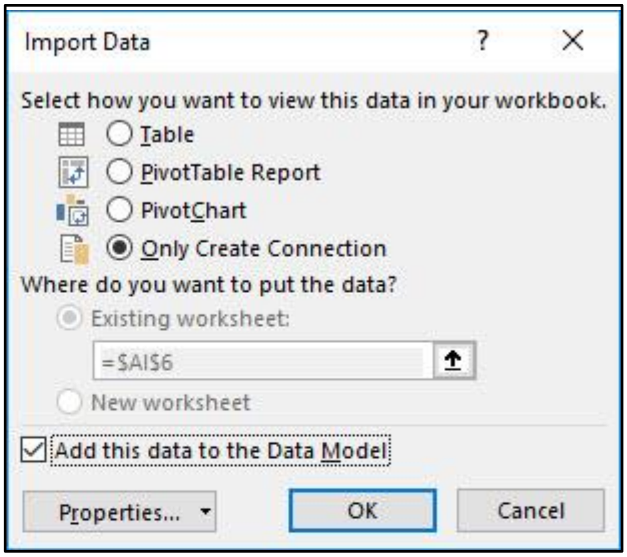
- In the next step, click OK, as seen here:



Encryption Support

We were unable to connect to the data source using an encrypted connection. To access this data source using an unencrypted connection, click OK.

- After using the Navigation window to select tables, then use the Power Query Editor to clean and transform the tables, Load the Tables as “Only Create Connection” and “Add this to the Data Model”, as seen here:



Import Data

Select how you want to view this data in your workbook.

○ Table

○ PivotTable Report

○ PivotChart

● Only Create Connection

Where do you want to put the data?

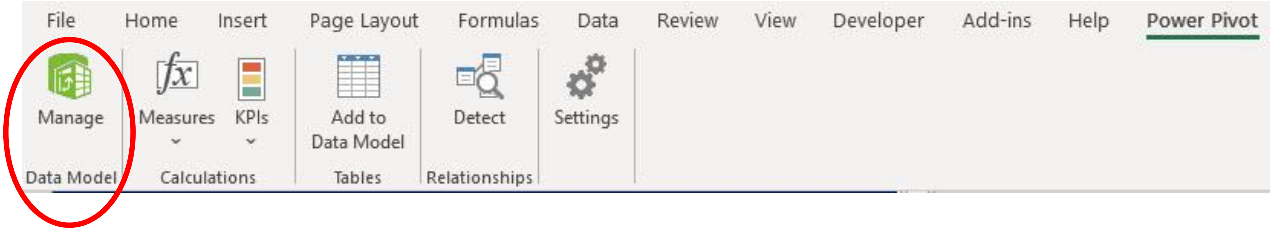
● Existing worksheet:

=AI6

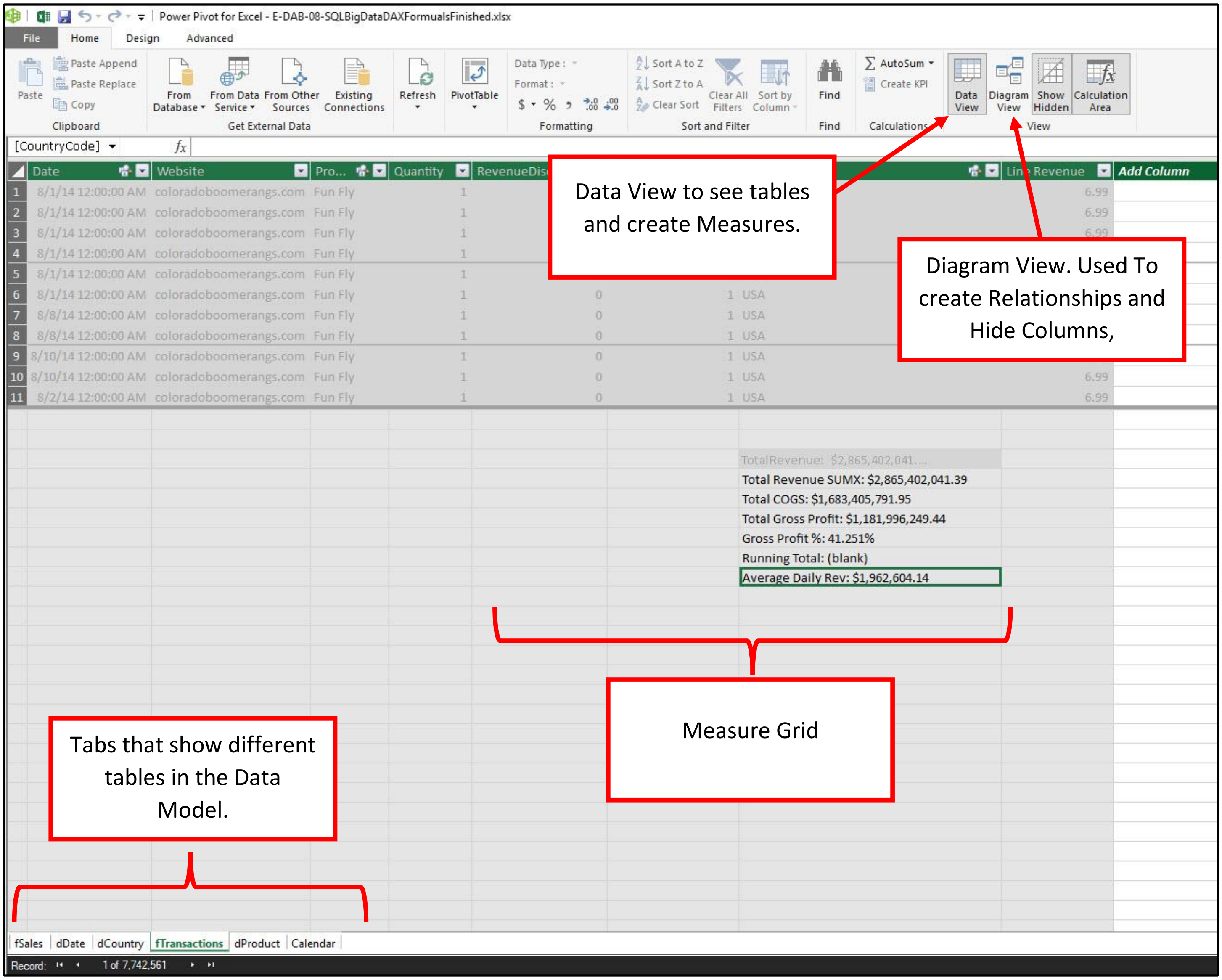
○ New worksheet

☑ Add this data to the Data Model

6. **Steps in Power Pivot Data Modeling**:
    1. Import Tables into the Columnar Database
    2. Create Date Table.
    3. Create Relationships between Tables
    4. Hide Columns, Tables and Measures that you do not want to see in the PivotTable Reporting Area.
    5. Create DAX Formulas
7. **Power Pivot Ribbon tab in Excel looks like this** :
    - Use "Manage" button in Data Model group to work on the Data Model in the "Power Pivot for Excel" Window.



8. **Power Pivot for Excel" Window looks like this** :



	Date	Website	Pro...	Quantity	RevenueDis			Lin Revenue	Add Column
1	8/1/14 12:00:00 AM	coloradoboomerangs.com	Fun Fly	1				6.99	
2	8/1/14 12:00:00 AM	coloradoboomerangs.com	Fun Fly	1				6.99	
3	8/1/14 12:00:00 AM	coloradoboomerangs.com	Fun Fly	1				6.99	
4	8/1/14 12:00:00 AM	coloradoboomerangs.com	Fun Fly	1					
5	8/1/14 12:00:00 AM	coloradoboomerangs.com	Fun Fly	1					
6	8/1/14 12:00:00 AM	coloradoboomerangs.com	Fun Fly	1	0	1	USA		
7	8/8/14 12:00:00 AM	coloradoboomerangs.com	Fun Fly	1	0	1	USA		
8	8/8/14 12:00:00 AM	coloradoboomerangs.com	Fun Fly	1	0	1	USA		
9	8/10/14 12:00:00 AM	coloradoboomerangs.com	Fun Fly	1	0	1	USA		
10	8/10/14 12:00:00 AM	coloradoboomerangs.com	Fun Fly	1	0	1	USA	6.99	
11	8/2/14 12:00:00 AM	coloradoboomerangs.com	Fun Fly	1	0	1	USA	6.99	

TotalRevenue: $2,865,402,041....
Total Revenue SUMX: $2,865,402,041.39
Total COGS: $1,683,405,791.95
Total Gross Profit: $1,181,996,249.44
Gross Profit %: 41.251%
Running Total: (blank)
Average Daily Rev: $1,962,604.14

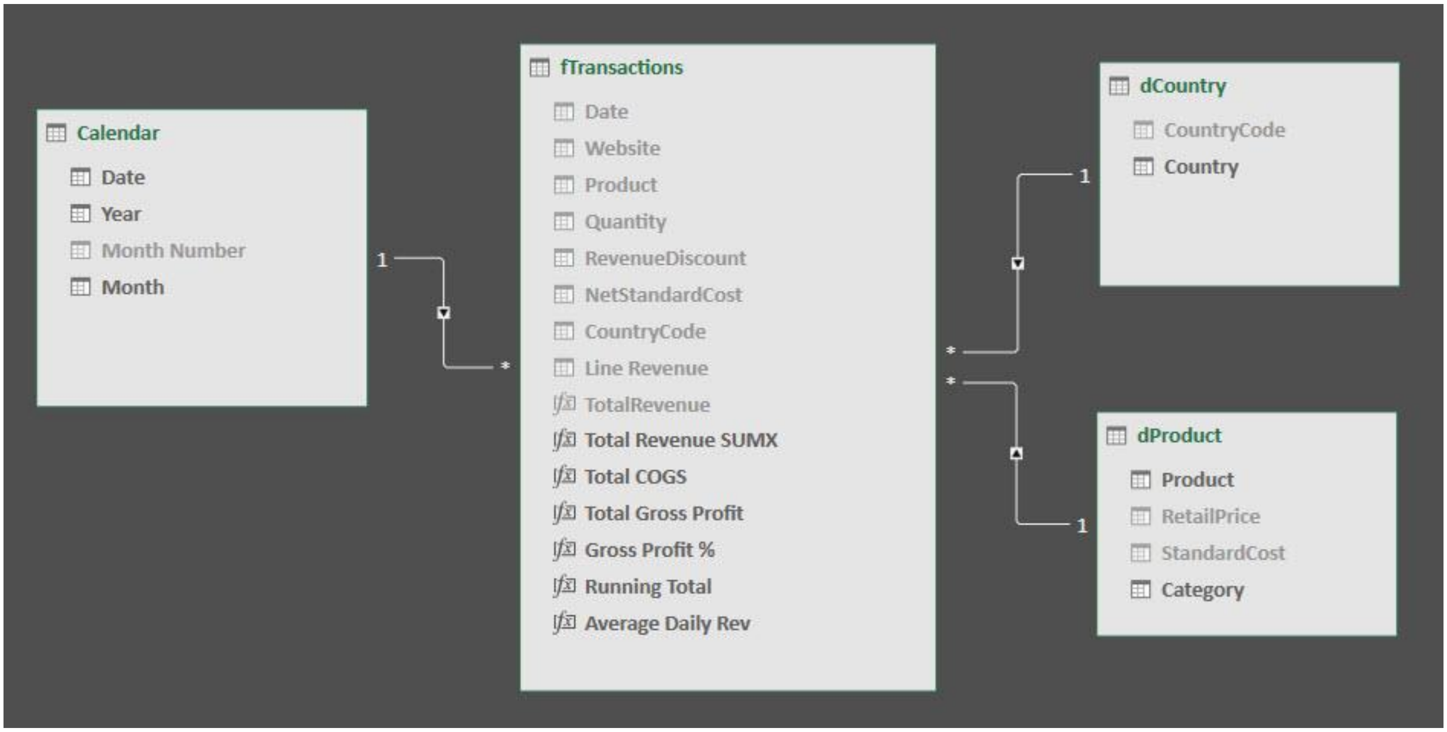
Data View to see tables and create Measures.

Diagram View. Used To create Relationships and Hide Columns,

Measure Grid

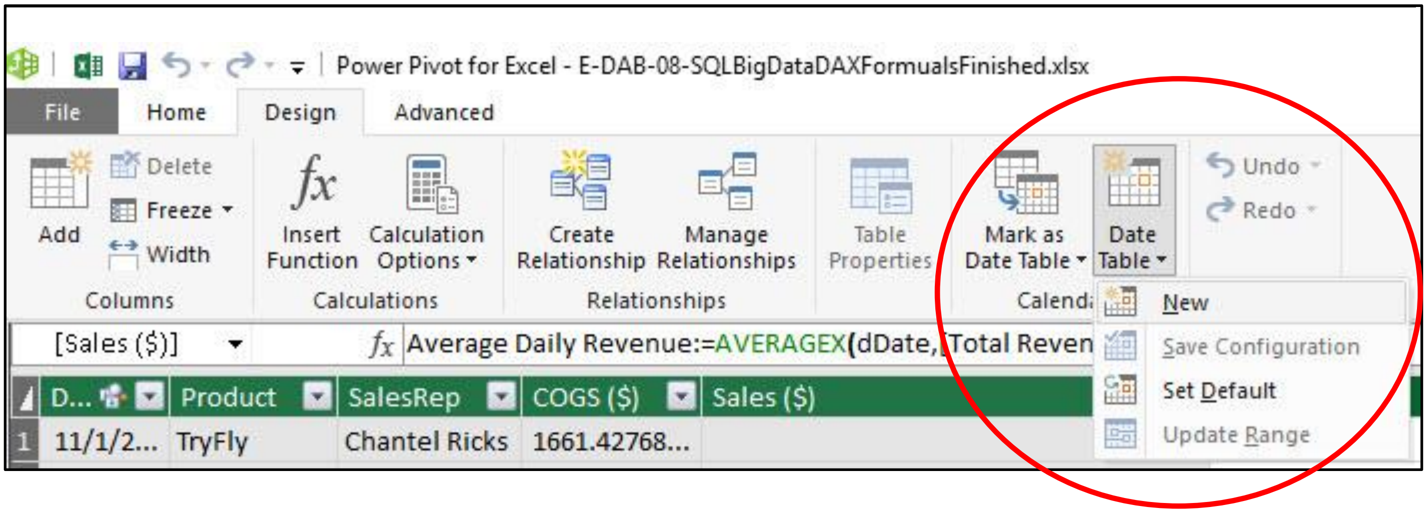
Tabs that show different tables in the Data Model.

9. **Diagram View** is where we can drag and drop Fields to create Relationships and hide Fields with the “Hide From Client Tools” feature:



10. **Creating a Date / Calendar Table**

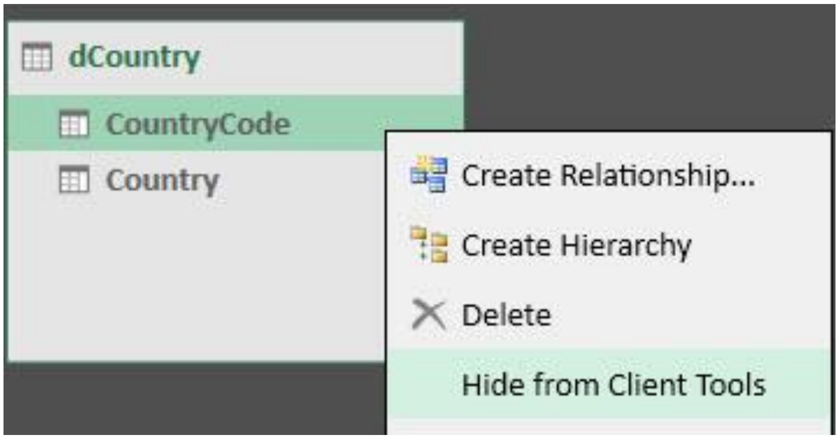
- Date Tables must have all the days in each year that is present in the Fact Table.
- Date Tables are Dimension Tables.
- Data Tables are required for the DAX Time Intelligence functions, like TOTALYTD.
- Attribute Columns like Month Name will sort alphabetically rather than by the calendar year, unless you use the Sort By feature and sort the column by an appropriate number column (for Month Name, sort by Month Number).
- To Create an automatic Data Table in Power Pivot:
    i. In the Excel Power Pivot window, click in one cell in the Fact Table.
    ii. In the Design Ribbon Tab, in the Calendar group, click the Date Table dropdown arrow, than click on New, as seen below:



```
Average Daily Revenue:=AVERAGEX(dDate,[Total Reven
```

D...	Product	SalesRep	COGS ($)	Sales ($)
11/1/2...	TryFly	Chantel Ricks	1661.42768...	

11. **Hide Column From Client Tool** :
    - In Design View you can right click an element such as a Table, Measure or Field / Column, and click on "Hide From Client Tools", as seen here:



12. **Reasons to use DAX Formulas in this video** :
    - Create One Formulas, with Number Formatting Attached, and then use it over and over.
    - Shows up in the PivotTable Field List in the Fact Table as a Formula.
    - Can do calculations like Average Daily Revenue that are harder to do in a Standard PivotTable.
13. **Why the name Power Pivot?**
    - Because Microsoft wanted to use the same amazing PivotTable user interface to drag and drop fields to make reports but with more Power.
    - The "Power" part of the name means:
        i. We can make PivotTables from "Big Data"
        ii. We can make PivotTables from multiple Tables
        iii. We can use DAX Formulas, which can process over big data efficiently and which allows us more varied calculations than in a Standard PivotTable.
    - The "Pivot" part of the name means we can use a PivotTable user interface, that we all know and love!

14. **Power Pivot & Power BI Data Model contains** :

- Columnar Database :
    - i. In RAM Memory Database. It is loaded each time you open the file.
    - ii. Efficient storage location for tables that compresses raw data and stores each column from the table separately as a unique list.
    - iii. Helps to reduce file size.
    - iv. Works with DAX Formulas to perform calculations quickly on Big Data.
- Relationships :
    - i. One-To-Many Relationships to allow multiple tables in PivotTable Field List.
    - ii. When we drag and drop a Field for the Row, Column and Filter areas of a report, we drag the Field from the One-Side. This leads to more efficient performance and helps to avoid errors if there are unmatched items in a relationship.
- DAX Formulas :
    - i. DAX = **D**ata **A**nalysis e**X**pressions
        1. "Expressions" is a synonym for "Formula"
    - ii. Calculated Columns.
        1. Used for Intermediate Calculations, like Revenue.
        2. Used to create attributes in Dimension tables, like Month and Year for dates.
        3. Use "Row Context" to access the values in each row of a table to make a calculation Row-By-Row.
    - iii. Measures
        1. Measure = Formula used in the Values Area of a PivotTable or in a Power BI Visualization.
        2. Implicit Measure = Formula that Power Pivot or Power BI creates automatically. Only use on small data sets when you are making simple calculations like sum of revenue.
        3. Explicit Measure = Formula you author. Preferred method for making calculations in a Data Model PivotTable Values Area or in a Power BI Visualization.
        4. Measures use "Filter Context" to see the Conditions and Criteria from the Row, Column, Filter and Slicer Areas of the PivotTable to filter the underlying tables and then make the calculation on only the records that match the conditions and criteria coming from the Row, Column, Filter and Slicer Areas of the PivotTable.
    - iv. Calculated Column or Measure?
        1. Calculated Columns:
            - a. Increase the use of RAM memory.
            - b. Are Calculated when you refresh or load the data.
        2. Measures:
            - a. Do NOT increase the use of RAM memory.
            - b. Are Calculated when you drop into Excel PivotTable or Power BI Visualization, or when conditions/criteria are changed in Row, Column, or Filter area of Report or Visualization.
        3. If Measure is calculating slowly each time you change condition / criteria in report or visualization, try converting calculation to a calculated column instead.

v. DAX Formula Syntax Conventions:

  1. Referring to Fields or Columns: Use Table Name and Field Name in Square Brackets for Columns.

     a. Like: **fTransactions[Line Revenue]** when referring to the “Line Revenue” Field from the fTransactions Table.

  2. Referring to Measures: Use the Measure Name in Square Brackets.

     a. Like **[Total Revenue]** when referring to the Measure “Total Revenue”.

vi. DAX Iterator Functions:

  1. DAX Iterator Function like SUMX and AVERAGEX can simulate what a Calculated Column can do.

  2. Iterator functions can take a formula and iterator across a table to calculate the intermediate values, and then use those values in an aggregate calculation like adding (SUMX) or averaging (AVERGAEX).

vii. We want to create our Measures in the Fact Table and then hide all the columns in the Fact Table. This forces the user to drag and drop fields from the Dimension Tables and leads to efficient performance and fewer errors.

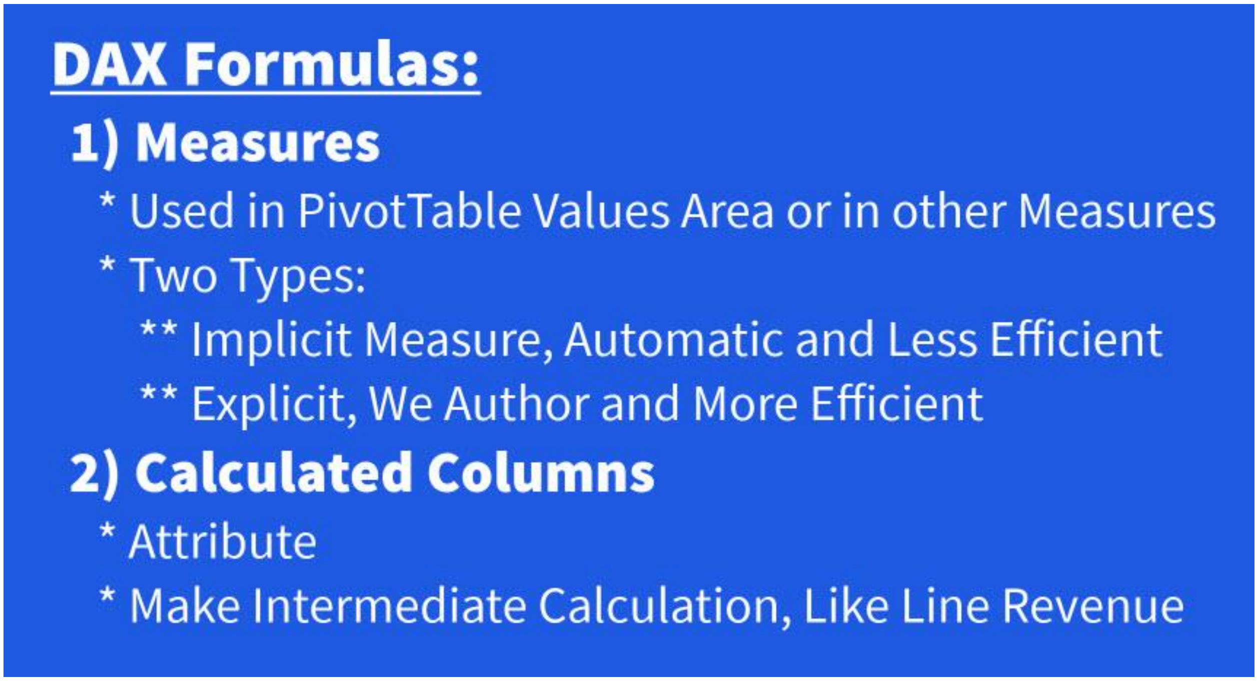
15. Filter Context for simple DAX formula:

i. When you drop a Measure into Values Area of PivotTable or into a Visualization, the Conditions / Criteria from the Row, Column and Filter Areas flow into the Data Model and filter the underlying tables so that show only the records that match the conditions and criteria. This is an internal mechanism of the Data Model that helps DAX formulas to calculate efficiently on Big Data.

ii. You can use the CALCULATE function to change the Filter Context. There are also functions like TOTALYTD that can change the Filter Context.

iii. There is much more to how the Filter Context and the CALCULATE Function works, but that is discussed in the next class, the Advanced Data Analysis Class, Microsoft Power Tools for Data Analysis (MSPTDA).

16. Summary of Types of DAX Formulas from Video :



DAX Formulas:

1) Measures

* Used in PivotTable Values Area or in other Measures

* Two Types:

** Implicit Measure, Automatic and Less Efficient

** Explicit, We Author and More Efficient

2) Calculated Columns

* Attribute

* Make Intermediate Calculation, Like Line Revenue

17. Summary of DAX Formulas used in Video :

[1] Total Revenue SUMX :=

```
SUMX(fTransactions,ROUND(RELATED(dProduct[RetailPrice])*fTransactions[Quantity]*(1-fTransactions[RevenueDiscount]),2))
```

[2] Total COGS :=

```
SUMX(fTransactions,ROUND(RELATED(dProduct[StandardCost])*fTransactions[Quantity]*fTransactions[NetStandardCost],2))
```

[3] Total Gross Profit :=

```
[Total Revenue SUMX]-[Total COGS]
```

[4] Gross Profit % :=

```
DIVIDE([Total Gross Profit],[Total Revenue SUMX])
```

[5] Running Total :=

```
TOTALYTD([Total Revenue SUMX],'Calendar'[Date])
```

[6] Average Daily Rev :=

```
AVERAGEX('Calendar',[Total Revenue SUMX])
```