



Data Analysis

Agenda

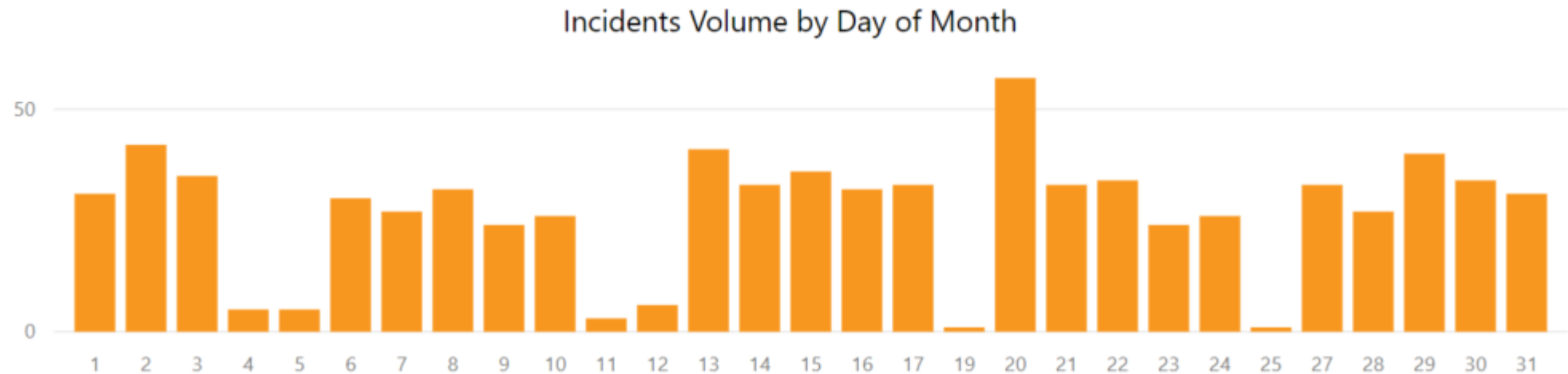
- What to measure
 - ❑ Queries Fundamentals
 - ❑ Data Modeling
 - ❑ PBIX Project File
 - ❑ Parameters & Templates

Data Discovery

- Data is siloed in many organization
 - Databases – OLAP/OLTP
 - Files – Text files, CSVs, Excel
 - Collaboration sites – SharePoint
 - Web Data
 - Social Media
 - Software As a Service data
- Bring them together to tell a story, providing 360° view of the environment

Measuring

- Geared toward 1 or 2 questions
 - Financial (Revenue, Expenses, Profits etc.)
 - Crime levels
 - Organization Health



Understanding Data

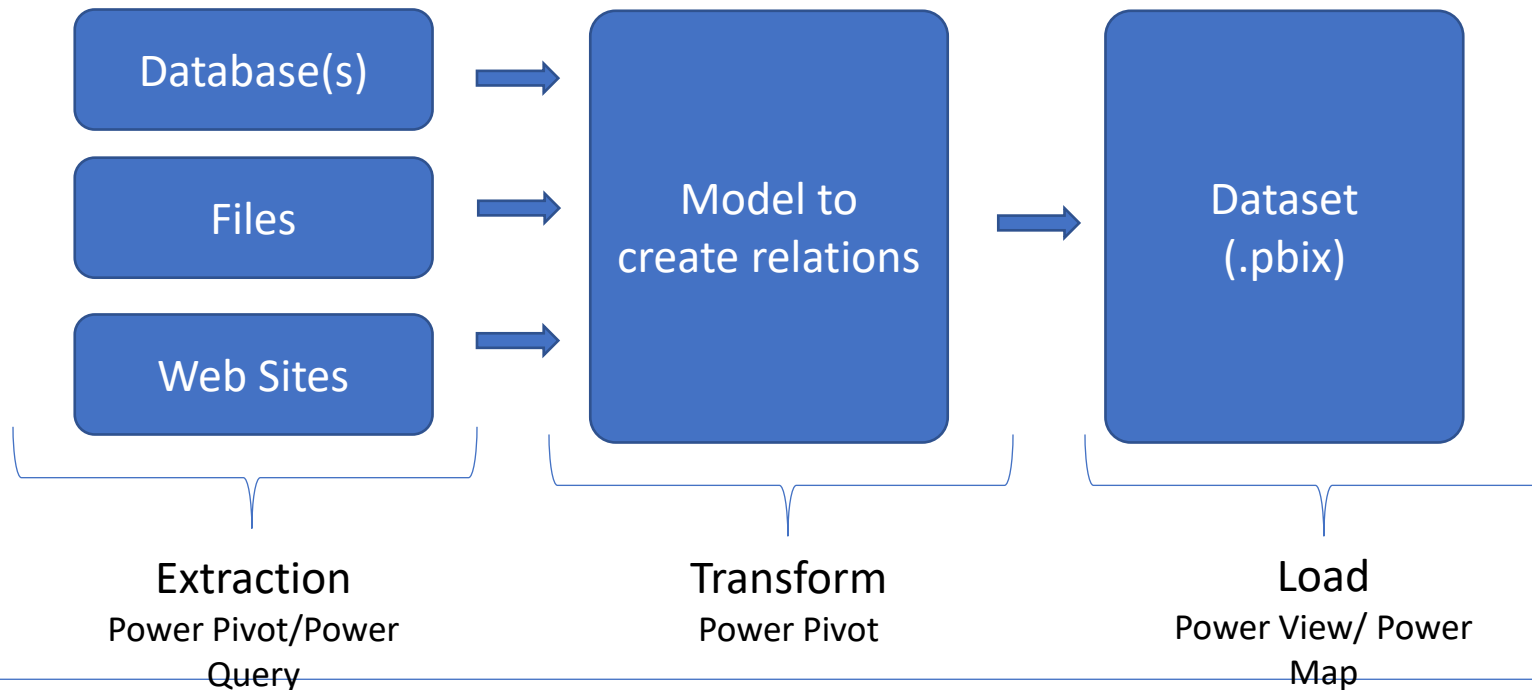
- Before Design
 - Understand your data
 - Define your grain – lowest level of details to measure
 - The lower the grain, the higher the data volume
 - Examples
 - Course enrollment? – measure by demographics
 - Department performance? – measure by area (No of Employees, Type of Work, Region)
 - Factors leading to performance decrease

Agenda

- ✓ What to measure
- Queries Fundamentals
- ☐ Data Modeling
- ☐ PBIX Project File
- ☐ Parameters & Templates

Data Exploration

- Power BI allows you to explore data
 - Data Extraction
 - Data Transformation
 - Loading



Transformations

- Provides integrated data manipulation platform – Query Editor
 - Very Powerful incorporation of Power Query
 - Only available to Power BI Desktop, also Excel has it.
 - Meant to shape and transform data before loading
 - Shows history of steps taken in data shaping and transformation

Query Editor Window

The screenshot displays the Power BI Query Editor interface. The top ribbon includes 'File', 'Home', 'Insert', 'Modeling', 'View', and 'Help'. The 'Home' ribbon is active, showing options like 'Get data', 'Excel', 'Power BI', 'SQL Server', 'Enter data', and 'Recent sources'. A red box highlights the 'Transform data' button, which has a dropdown menu containing 'Transform data' and 'Data source settings'. Below this, the 'Visualizations' pane is visible on the right. The main area shows a query named 'AI Revenue by market' with a table of data. The table has columns: 'date', 'day', 'month', 'FirstOfMonth', 'MonthName', and 'week'. The data rows show dates from 1/1/2018 to 1/10/2018, with corresponding day, month, and week values.

File Home Insert Modeling View Help

Paste Cut Copy Format painter Clipboard

Get data Excel Power BI SQL Server Enter data Recent sources Data

Transform data Refresh New visual Text box More visuals Insert Calculations Share Visualizations

Transform data Data source settings Edit parameters

AI Revenue by market

File Home Transform Add Column View Tools Help

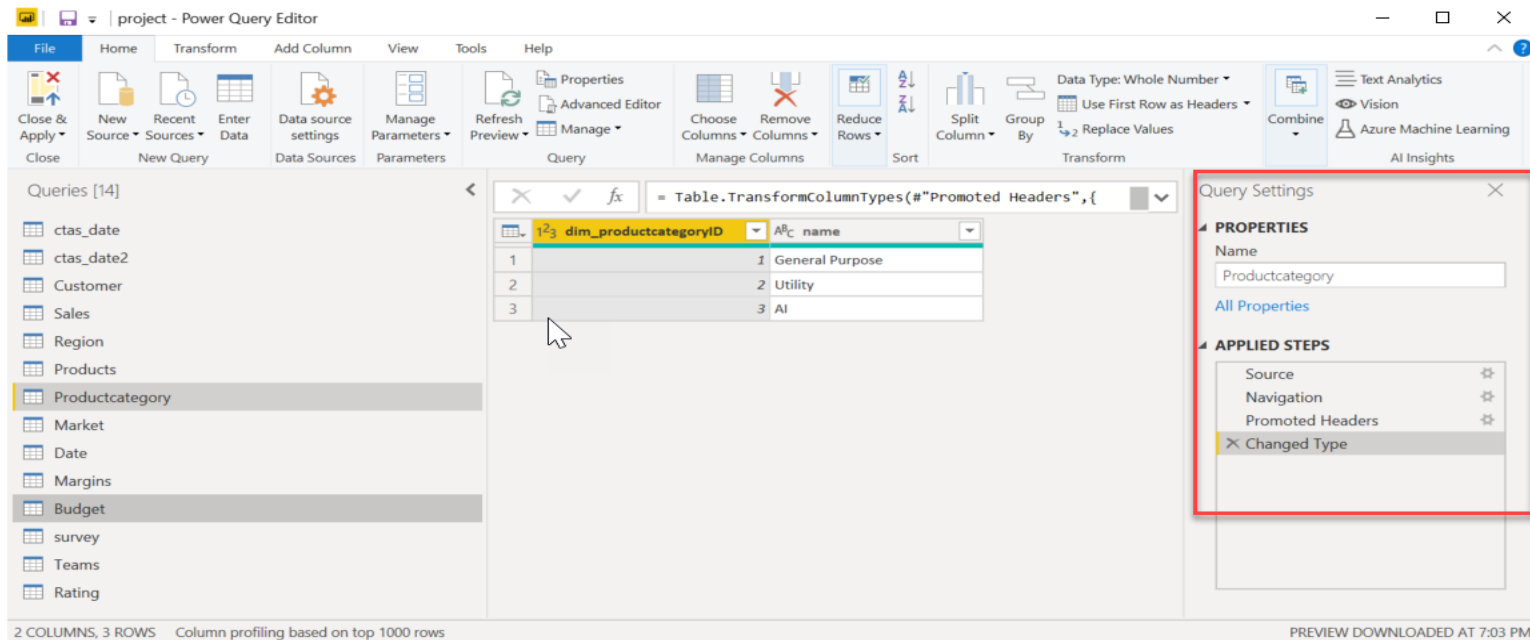
Close & apply Close New Source Recent Sources Enter Data Data source settings Manage Parameters Refresh Preview Advanced Editor Query Choose Columns Remove Columns Keep Rows Remove Rows Sort Split Column Group By Data Type: Date Use First Row as Headers Replace Values Merge Queries Append Queries Combine Files Text Analytics Vision Azure Machine Learning AI Insights

Queries [14]

| | date | day | month | FirstOfMonth | MonthName | week |
|----|-----------|-----|-------|--------------|-----------|------|
| 1 | 1/1/2018 | | | | | |
| 2 | 1/2/2018 | | | | | |
| 3 | 1/3/2018 | | | | | |
| 4 | 1/4/2018 | | | | | |
| 5 | 1/5/2018 | | | | | |
| 6 | 1/6/2018 | | | | | |
| 7 | 1/7/2018 | | | | | |
| 8 | 1/8/2018 | | | | | |
| 9 | 1/9/2018 | | | | | |
| 10 | 1/10/2018 | | | | | |

History

- History of Transformations
 - Transformations are sequential
 - Steps have formula attached to them
 - They can be replayed

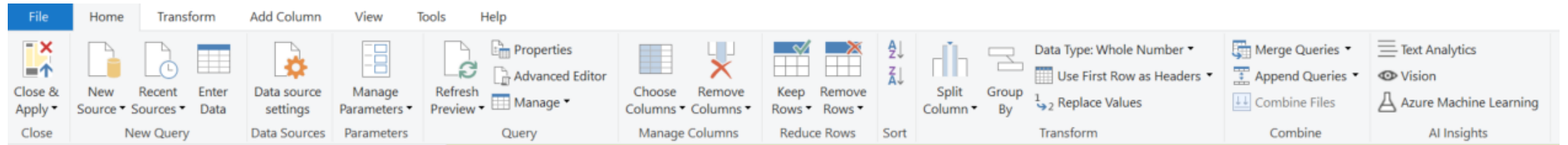


Query Editor Functions

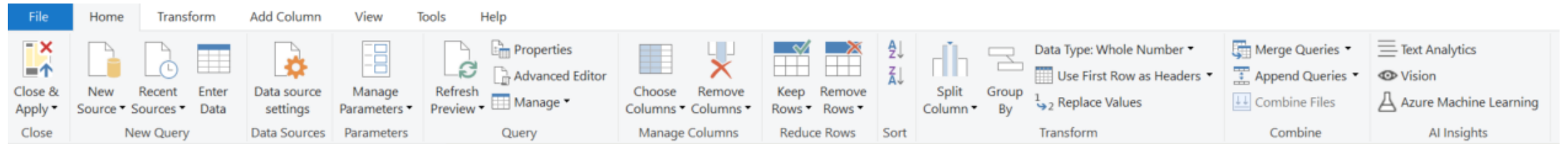
- Rename Columns
- Concatenate Columns
- Transpose data
- Pivot/Unpivot
- Split Columns
- Merging Columns
- Clean – remove non-printable characters

Tabs

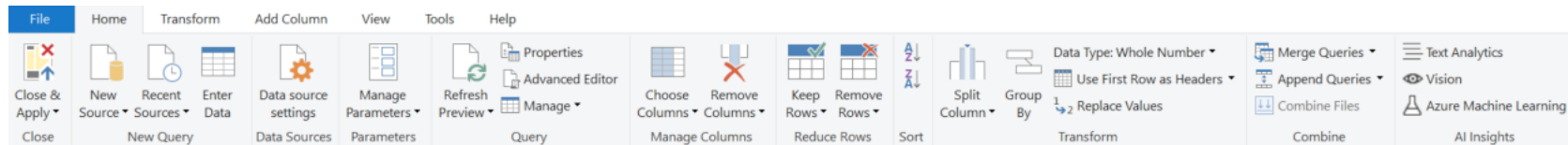
Home



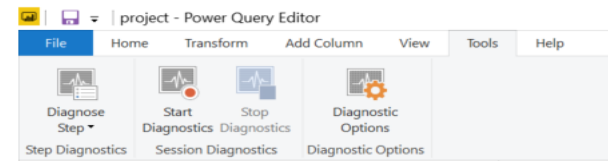
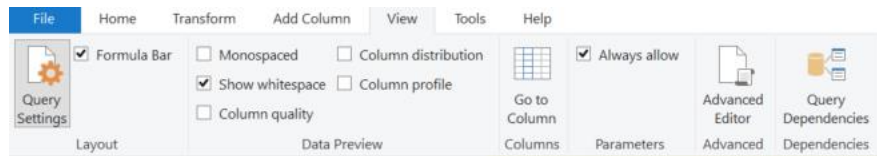
Transform



Add Columns

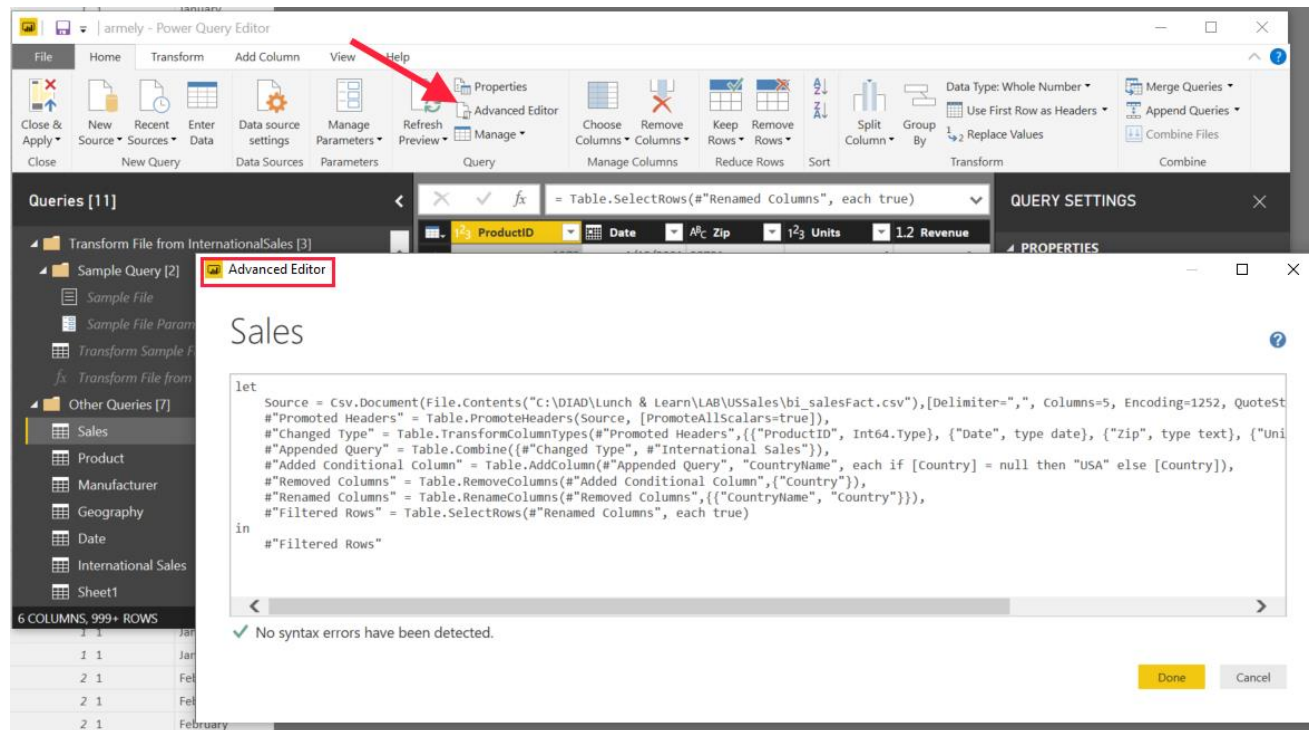


View & Tools



Advanced Editor

- Incorporates M-Query language, derived from F#
 - Query Editor generated M code behind the scene
 - Can edit the M-Code (Advanced Skills)



Data Quality

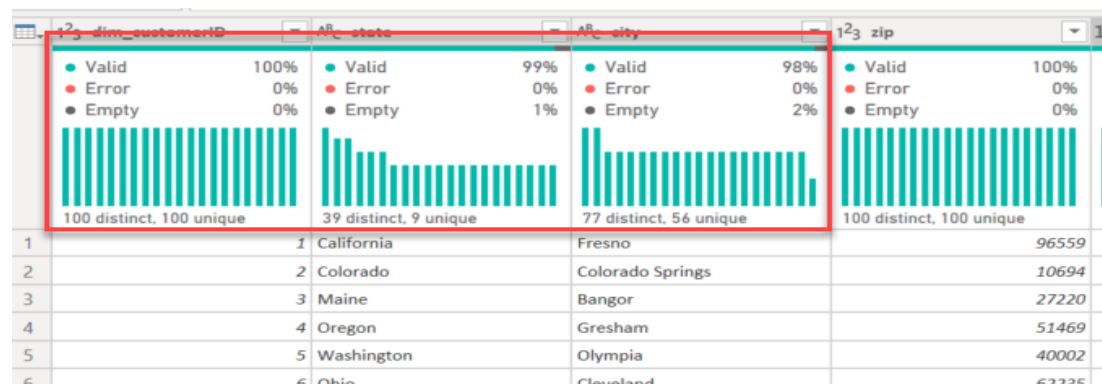
- Power BI offers a quick and easy way to spot data quality and statistics of the data.

formColumnTypes("#Promoted Headers",{{"dim_customerID", Int64.Type}, {"state", type text}, {"city", type text}, {"zip", Int64.Type}, {"latitude", type number}

| dim_customerID | state | city | zip | latitude | longitude | customer_name | address |
|----------------|------------|------------------|-------|-----------|-----------|-------------------------------|---------------------|
| 1 | California | Fresno | 96559 | 73.88975 | 19.98367 | Erat Incorporated | |
| 2 | Colorado | Colorado Springs | 10694 | -53.39357 | -47.61922 | Curabitur Dictum Industries | |
| 3 | Maine | Bangor | 27220 | -38.60279 | 93.12871 | Sed Hendrerit Company | |
| 4 | Oregon | Gresham | 51469 | -5.82802 | -17.64205 | Interdum Curabitur Foundation | |
| 5 | Washington | Olympia | 40002 | -21.45314 | 44.70458 | Mauris Erat LLC | |
| 6 | Ohio | Cleveland | 62235 | -83.9826 | 102.89121 | Ante Consulting | |
| 7 | Ohio | Gary | 49969 | -65.838 | -108.3297 | Quis Foundation | 495-3037 Semper St. |

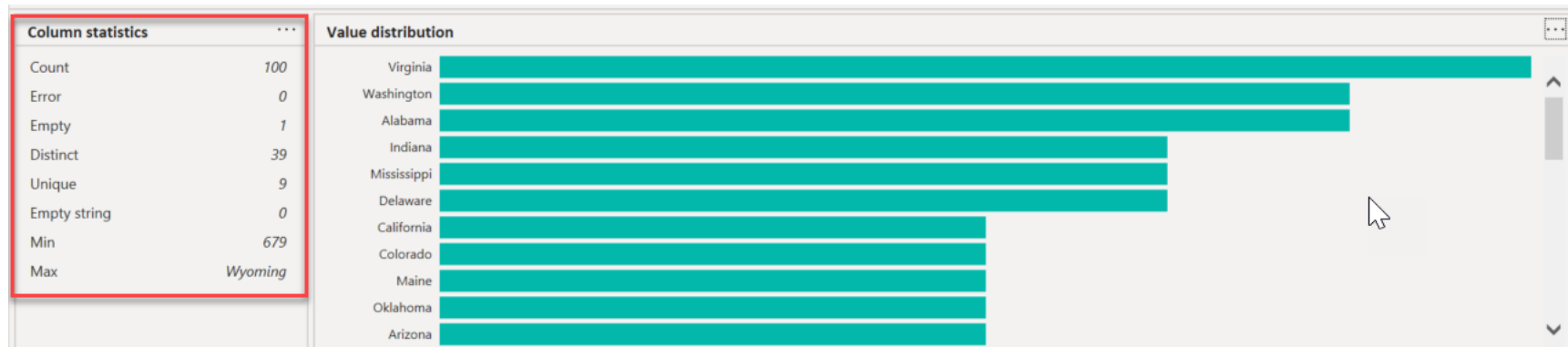
address
91 (91%) Valid | 0 (0%) Error | 9 (9%) Empty
Remove Empty ...

- Column Quality & Distribution



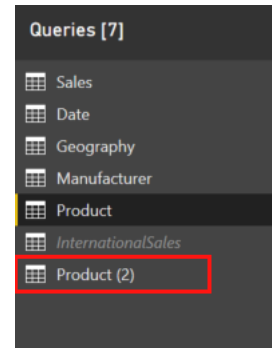
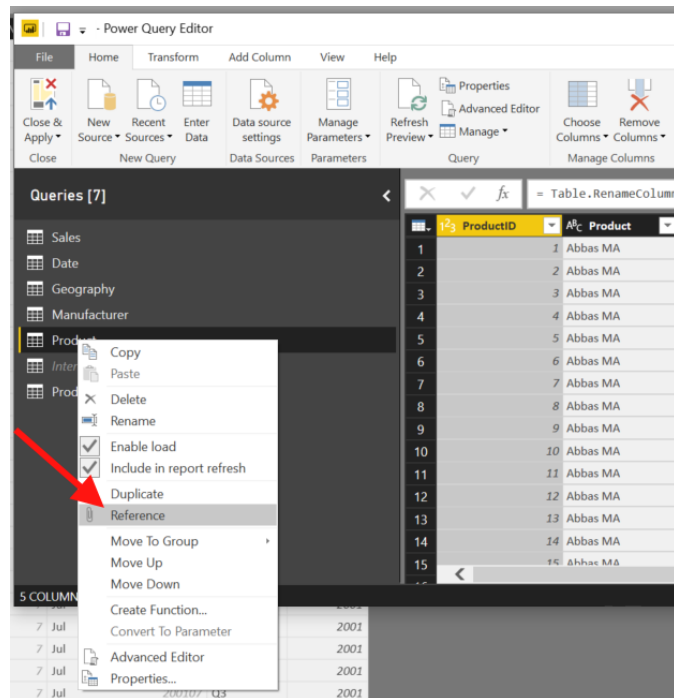
Data Quality Cont.

- Column Profile

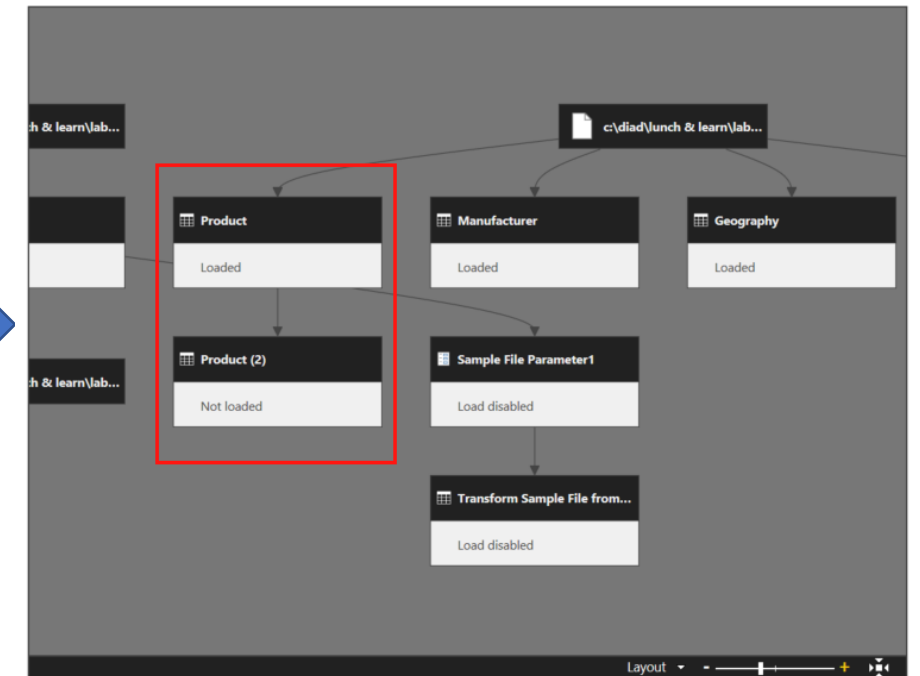


Queries

- Serves as source of other queries
 - Allow split to other queries
 - Reference can be used to create new query base

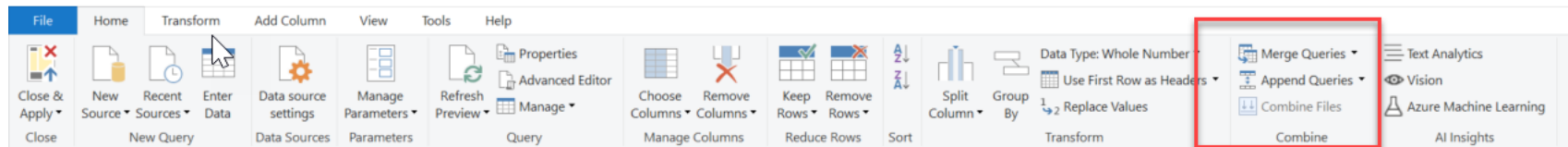


Query Dependencies



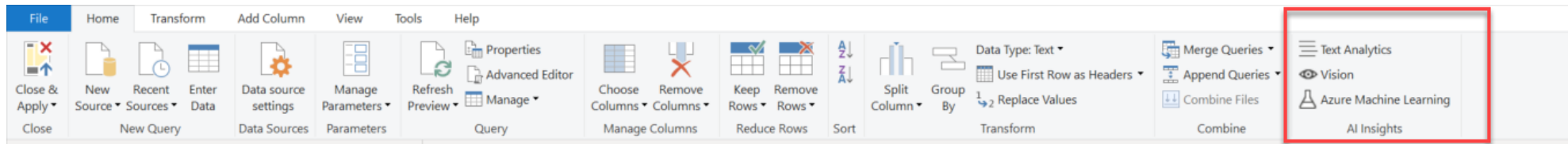
Queries Cont.

- Can be Merged or Appended
 - Merge allows combining columns from different tables
 - Append allows combining rows from different tables



AI Insights

- Power BI comes with a collection pre-trained machine learning models.
- Accessible from **Power Query Editor**



- Text Analytics - Sentiment Analysis, Key Phrase Extraction, Language Detection
- Vision – Image Tagging
- Azure Machine Learning – Azure ML Pre-Trained models, write your own model.



Data Transformations DEMO

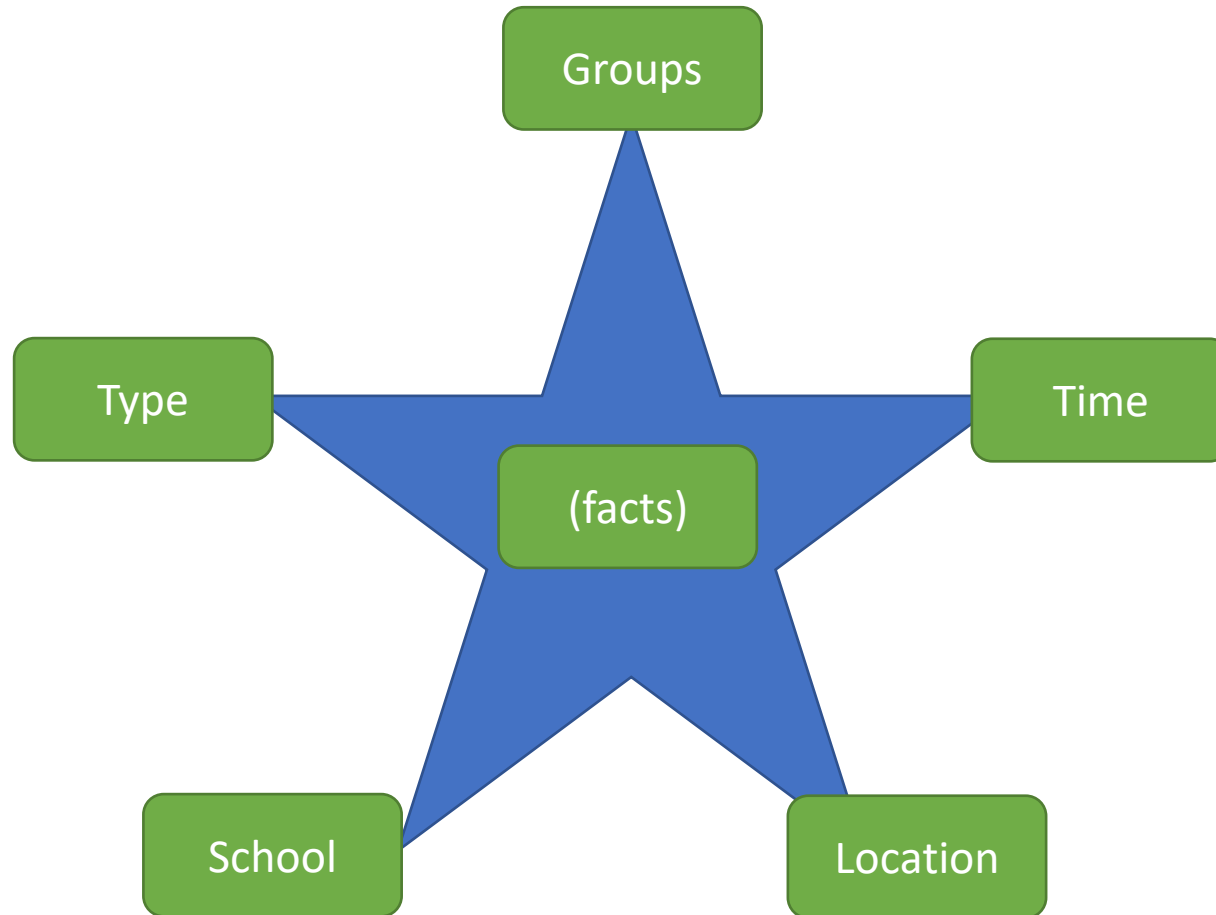
Agenda

- ✓ What to measure
- ✓ Queries Fundamentals
- Data Modeling
- ❑ PBIX Project File
- ❑ Parameters & Templates

Data Modeling

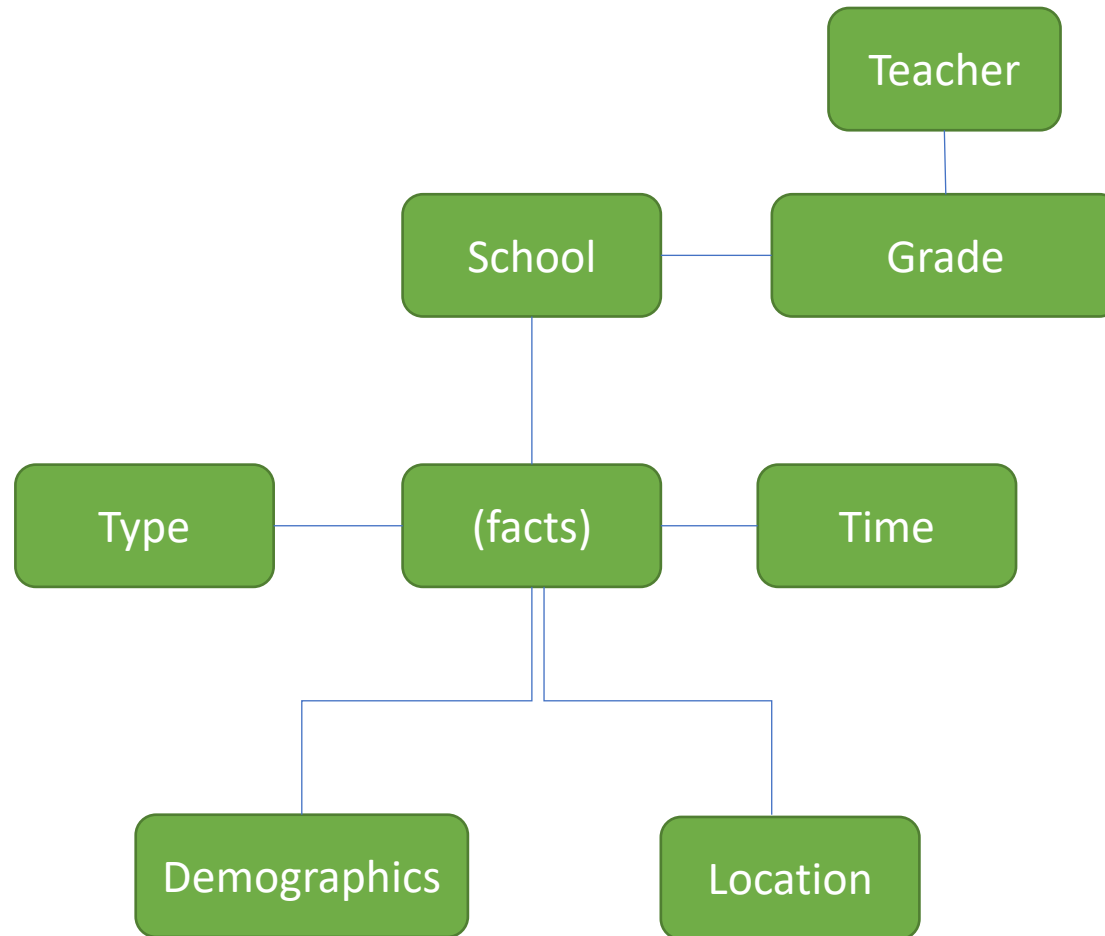
- Data modeling is the process of connecting multiple data sources to create a unique model that suites your needs
 - example (Date Table, Customers Table, Sales figures, Products, Regions)
- Model is enhanced by relationships
- Models are defined by
 - Dimensions – information about the data (strings & text)
 - Facts – Measurable data (10, 5, 20% etc..)
- Create relationship between fact at dimensions

Star Schema



- Easiest to Model
- Has 1:* relationship

Snowflake Schema



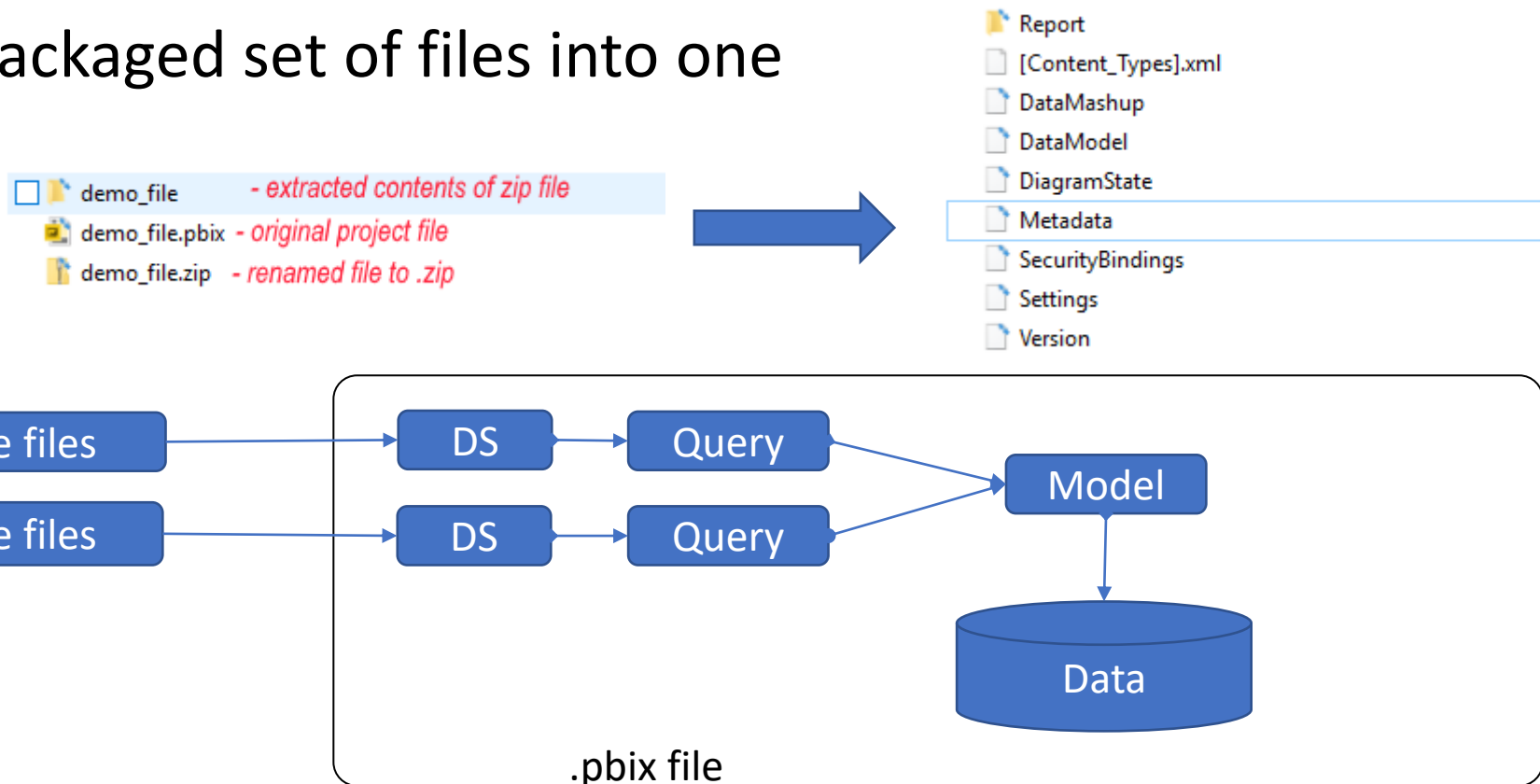
- Can be very Complex
- More Join are required
- Are flexible
- Improved performance

Agenda

- ✓ What to measure
- ✓ Queries Fundamentals
- ✓ Data Modeling
- PBIX Project File
- Parameters & Templates

PBIX Project File

- PBIX is the file extension used by Power BI desktop projects (<filename.pbix>)
- A packaged set of files into one



Agenda

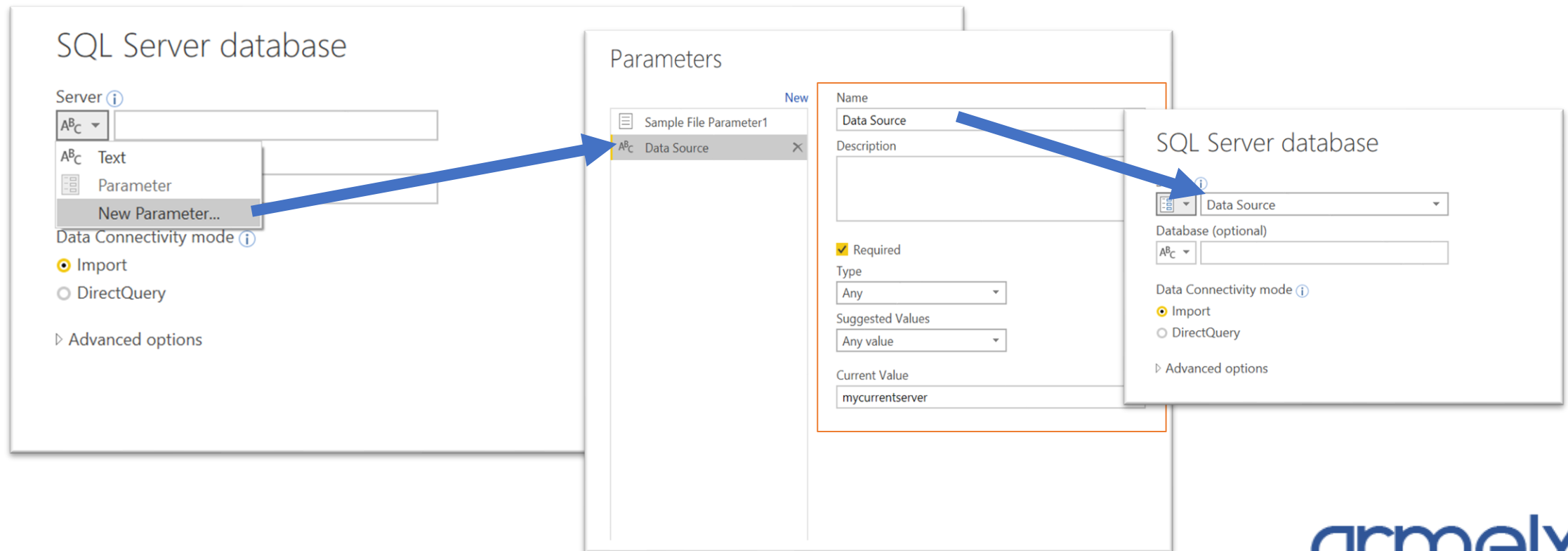
- ✓ What to measure
- ✓ Queries Fundamentals
- ✓ Data Modeling
- ✓ PBIX Project File
- Parameters & Templates

Parameters

- Query Parameters – a variable used to pass arguments between functions and procedures
 - Project level configurable settings
 - Referenced from DAX Queries
- Parameters in Power BI used in
 - Data Sources – ability to move from Development/QA/Production
 - Replace value
 - Remove rows
 - Filter Rows

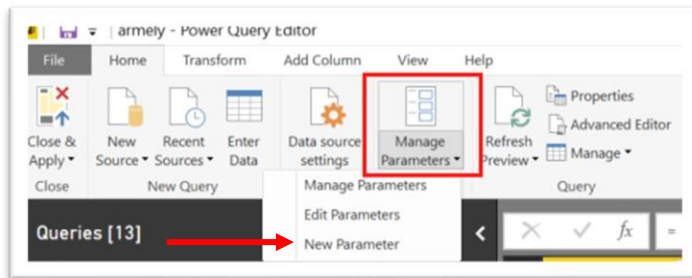
Data Source Parameters

- Used to change data sources – Mostly applicable in structure/Semi-Structured data



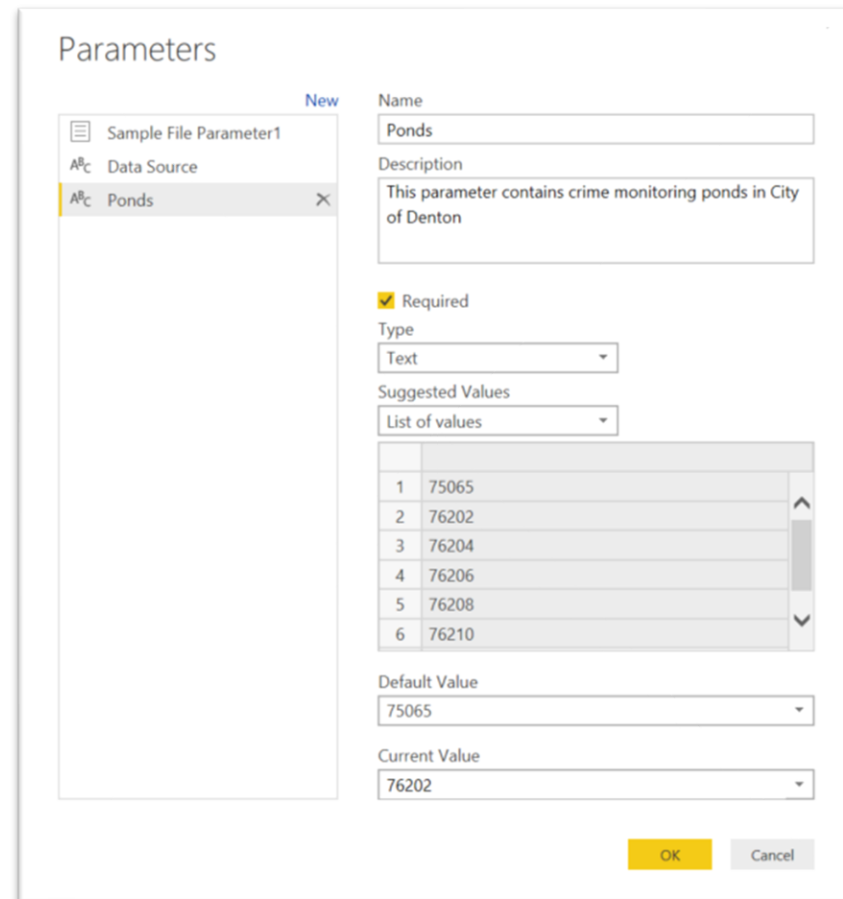
Query Parameters

- Query Parameter created using Manager Parameters Menu
 - Have to be in Query editor to use Query Parameters



- Properties

- Name
- Description
- Required
- Data Type
- Suggested values
- Default value
- Current value

A screenshot of the 'Parameters' dialog box. The dialog has a title bar 'Parameters' and a 'New' button. It shows a list of parameters on the left, including 'Sample File Parameter1' and 'Ponds'. The 'Ponds' parameter is selected. The right side of the dialog shows the properties for 'Ponds': Name (Ponds), Description (This parameter contains crime monitoring ponds in City of Denton), Required (checked), Type (Text), Suggested Values (List of values), Default Value (75065), and Current Value (76202). There are 'OK' and 'Cancel' buttons at the bottom right.

| 1 | 75065 |
|---|-------|
| 2 | 76202 |
| 3 | 76204 |
| 4 | 76206 |
| 5 | 76208 |
| 6 | 76210 |

Parameters Cont.

- Added as part of Queries
- Using Parameters in Queries

Add Conditional Column

Add a conditional column that is computed from the other columns or values.

New column name
HighCrime

| | Column Name | Operator | Value | | Output |
|----|-------------|----------|-------|------|-------------------------------|
| If | Zip | equals | Ponds | Then | ABC 123 "High Crime Area" ... |

Add rule

Otherwise

| | |
|---------|----------|
| ABC 123 | "Normal" |
|---------|----------|

Filter Rows

☐ Basic ☒ Advanced

Keep rows where "

| And/Or | Column | Operator | Value |
|--------|--------|----------|-----------------------------|
| | Zip | contains | Ponds |
| And | Zip | | ABC Enter or select a value |

Add Clause

What-if Parameter

- A variable that can be used to visualize and quantify key values in a report.
- What-if parameters allows the user to pass values to define an outcome.

What-if parameter

Name

Discount percentage

Data type

Decimal number

Minimum

0

Maximum

0.50

Increment

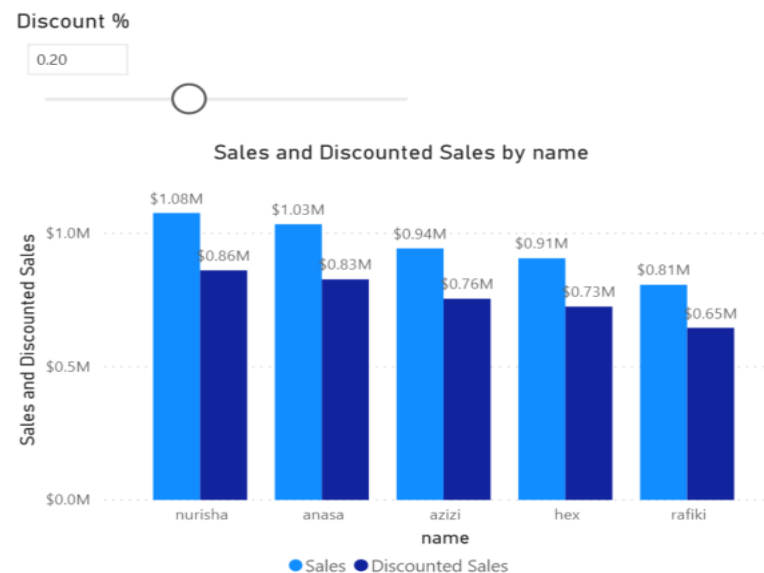
0.05

Default

☒ Add slicer to this page

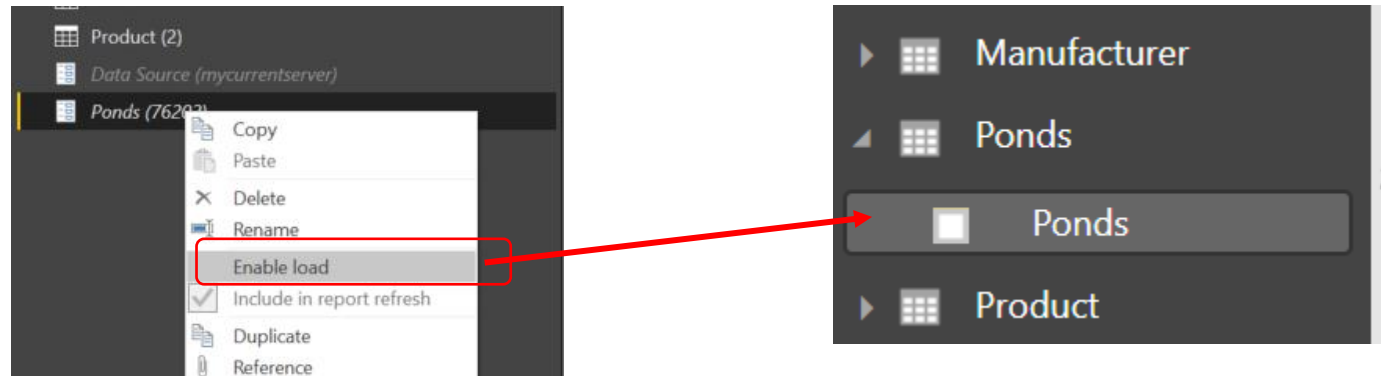
OK

Cancel



Part of Data Model

- Enabled Parameters becomes part of your datasets

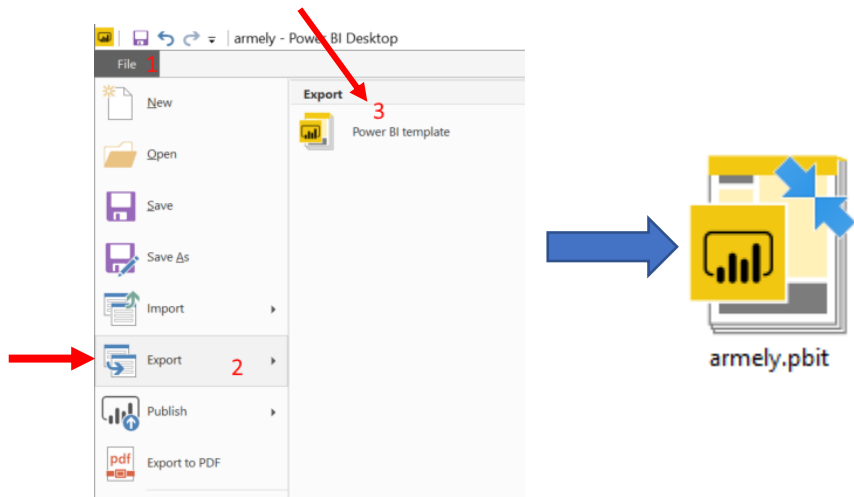




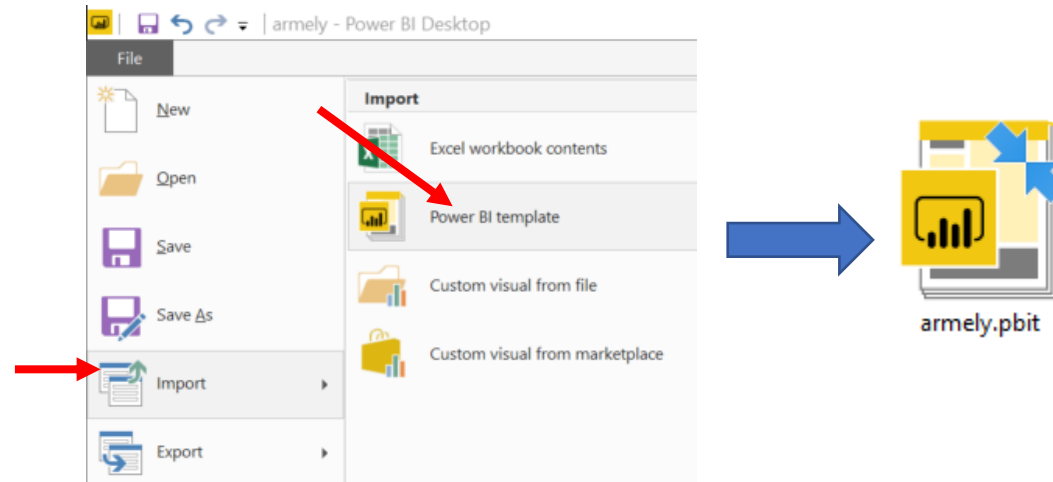
Parameters DEMO

Templates

- PBIX Project is exportable as a project template file
 - Template file created with file extension of PBIT
 - Template contains everything other than data
 - Can be imported to create new pbix projects



Export template Process



Import template Process



Templates DEMO

Agenda

- ✓ What to measure
- ✓ Queries Fundamentals
- ✓ Data Modeling
- ✓ PBIX Project File
- ✓ Parameters & Templates



Lab 2