Designing Queries to Extract and Transform Data

Agenda

➢ Deciding What To Measure

• Query Design Fundamentals

• Designing Data Model using a Star Schema

• Working with the Query Editor Window

• Importing Content From SharePoint Online

• Designing with Function Queries

• Understanding Parameters and Template Files

Data Discovery

- Data can live in a variety of sources
  - Files (e.g. CSV file, Excel workbook)
  - OLTP Databases
  - OLAP Databases
  - SharePoint Lists and Document Libraries
  - Azure-based services
  - Online services & SaaS applications

Deciding What To Measure

- You Must Determine Measurable Objectives
  - Financial (revenue, expenses, profit margin, etc.)
  - Business processes efficiency
  - Customer Satisfaction Levels

Defining Grain Statements

- Grain statements should be defined in initial design phase
    - Grain statements helps determine requirements for BI queries
    - Grain statements can be created & understood by business users

- Example grain statements for BI project at Wingtip Toys
    - What was the total sales revenue over the last 4 years?
    - What was the sales revenue by year, quarter and month?
    - What was the sales revenue by region, state, city and zip code?
    - What was the sales revenue by category, subcategory and product?
    - What was the growth in sales revenue from month to month in 2013?
    - What was profit margin for each product by year, quarter and month?
    - Have their been any products with significantly decreasing profit margin?

Agenda

- ✓ Deciding What To Measure
- ➢ Query Design Fundamentals
- • Designing Data Model using a Star Schema
- • Working with the Query Editor Window
- • Importing Content From SharePoint Online
- • Designing with Function Queries
- • Understanding Parameters and Template Files

Power BI Desktop is an ETL Tool

- ## ETL process is essential part of any BI Project
  - **Extract** the data from wherever it lives
  - **Transform** the shape of the data for better analysis
  - **Load** the data into dataset for analysis and reporting
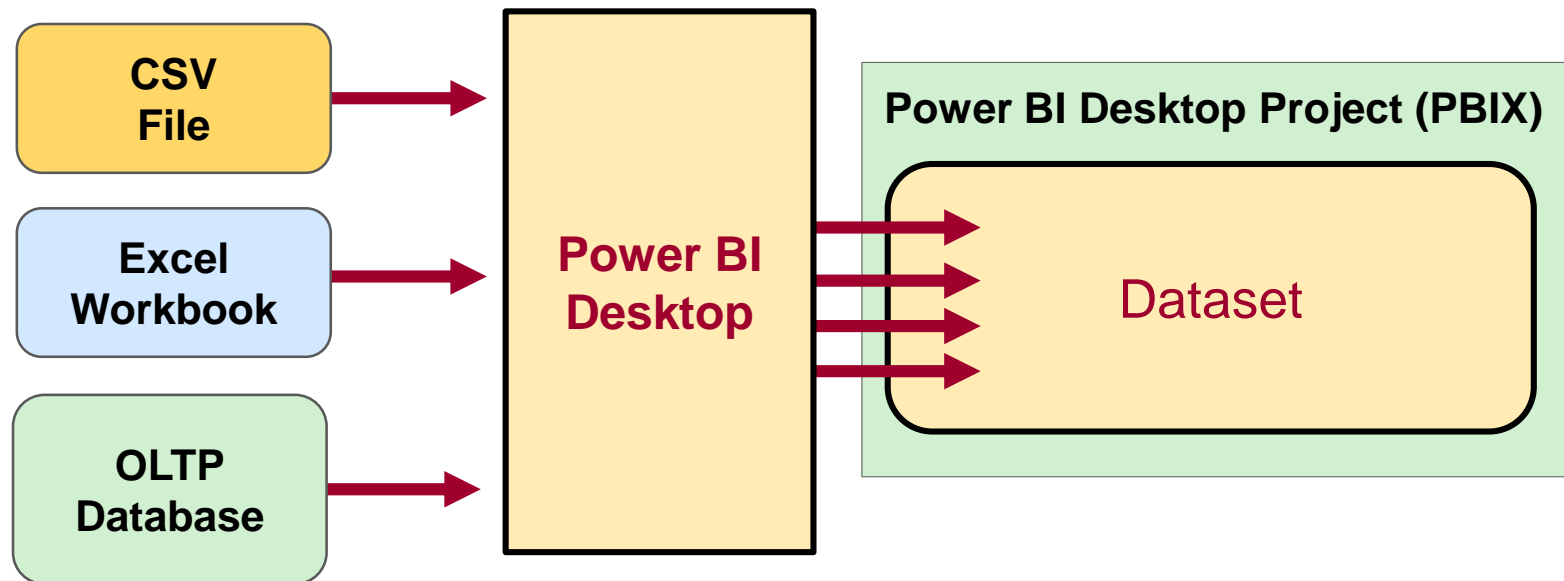
Query Editor Window

- Power BI Desktop provides separate Query Editor window
  - Provides powerful features for designing queries
  - Displays list of all queries in project on the left
  - Displays **Properties** and **Applied Steps** for selected query on right
  - Preview of table generated by query output shown in the middle
  - Query can be executed using **Apply** or **Close & Apply** command

Query Steps

- A query is created as a sequence of steps
  - Each step is a parameterized operation on the data
  - Each step has formula which can be viewed/edited in formula bar
  - Query starts with Source step to extract data from a data source
  - Additional steps added to perform transform operations on data
  - You can replay query operations one by one by clicking on steps

Advanced Editor

- Power BI Desktop based on "M" functional language
  - Query in Power BI Desktop saved as set of M statements in code
  - Query Editor generates code in M behind the scenes
  - Advanced users can view & modify query code in Advanced Editor

Working with Web Data Sources

- Many public websites publish data using HTML tables
 - Power BI desktop can scrape data from tables in HTML pages

Understanding Query Input and Output

- PBIX project is container for data sources and queries
  - Queries created and saved within scope of Power BI project
  - Queries can pull data from local files
  - Queries can pull data from external content sources
  - Queries main purpose is to load imported data into data model
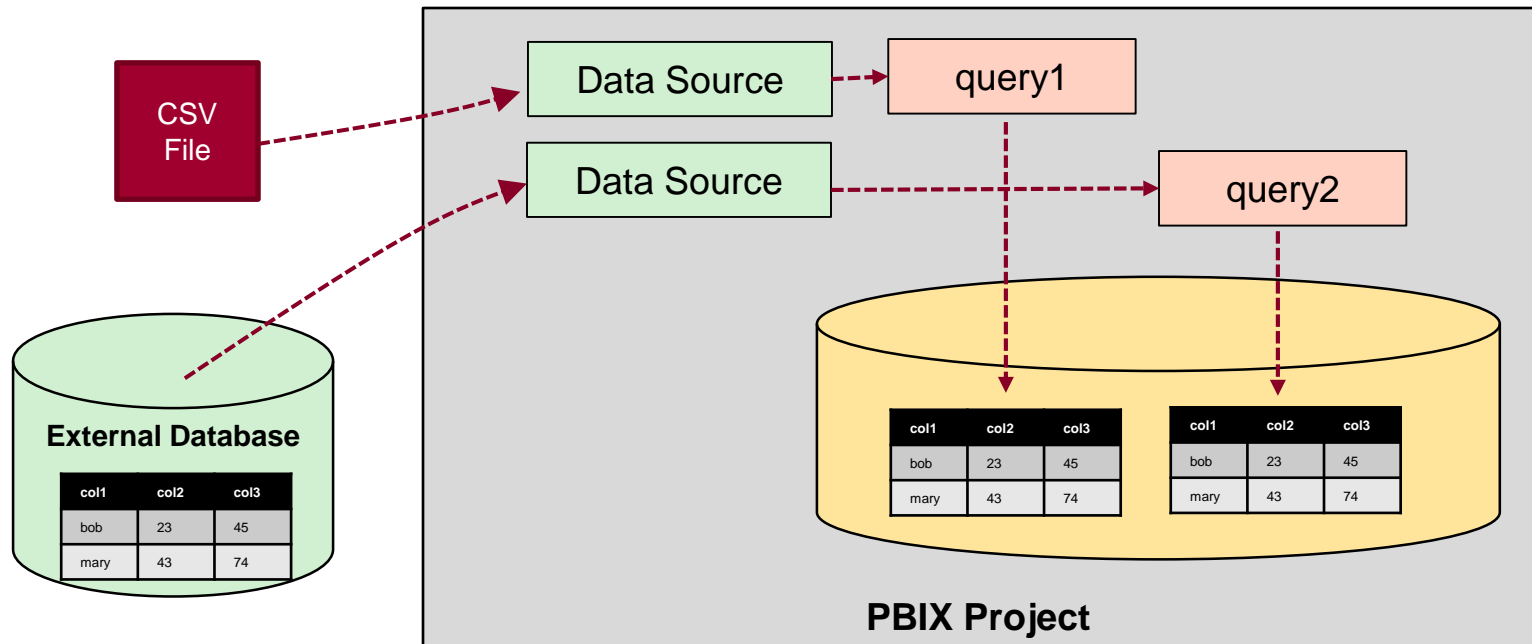
Query Composition

- Query can serve as source for other queries
  - Allows for creation of reusable base queries & query composition
  - Complexity can be hidden in base queries
  - **Reference** command creates new query based on another query
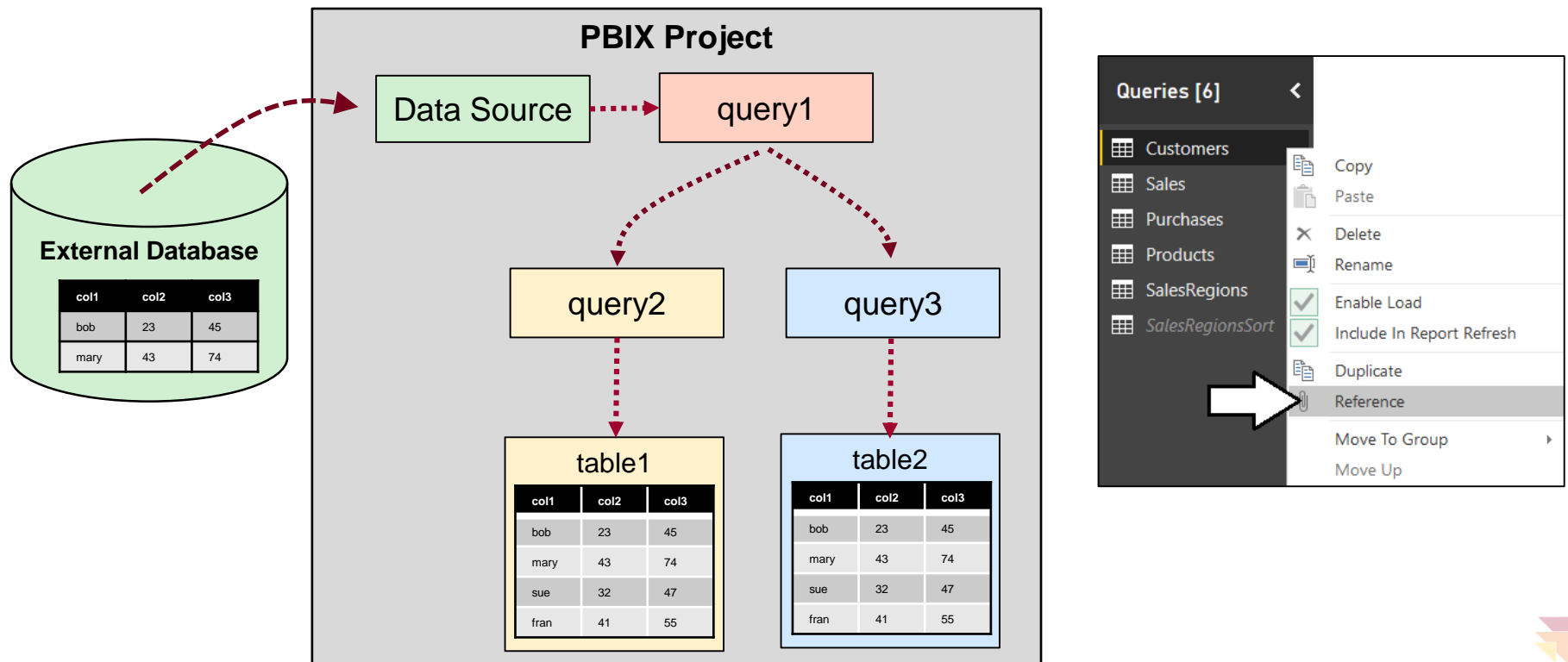
Combining Queries

- Query can be merged or appended with another query
  - Merge operation allows you combine columns from two tables
  - Append operation allows you to combine rows from two tables
- Two queries are combined into single output for loading
  - Load settings of main query determines where output is loaded
  - Secondary query acts as source for main query
  - Secondary query be can created with connection-only load setting
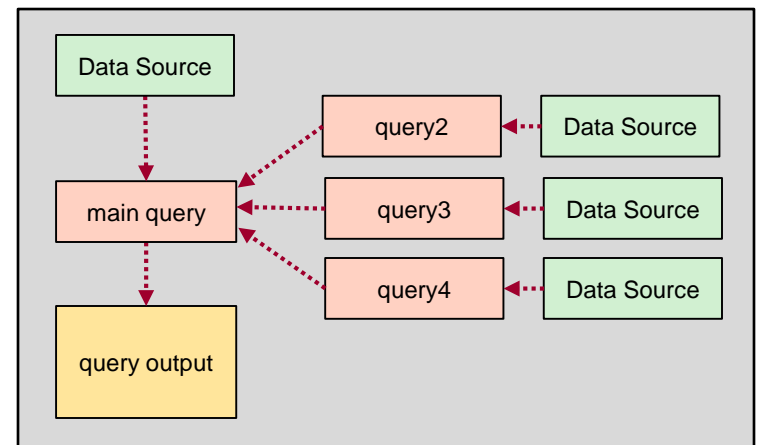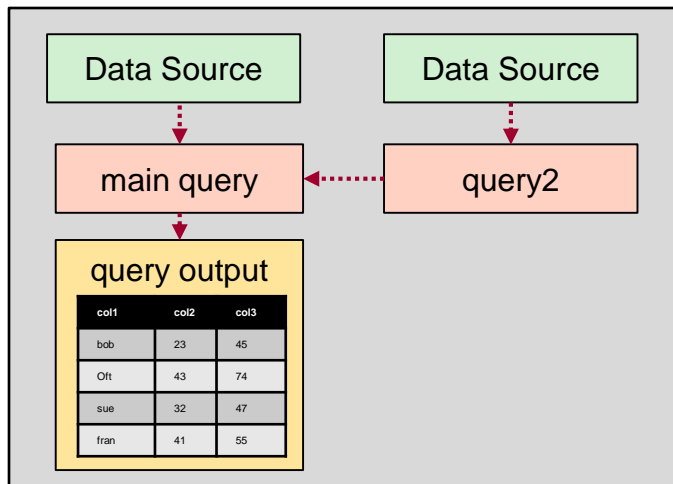
Agenda

- ✓ Deciding What To Measure
- ✓ Query Design Fundamentals
- ➢ Designing Data Model using a Star Schema
- • Working with the Query Editor Window
- • Importing Content From SharePoint Online
- • Designing with Function Queries
- • Understanding Parameters and Template Files

Sample OLTP Database: WingtipSalesDB

- Online Transaction Processing (OLTP) System
  - Used for real-time data access and transaction-based data entry
  - Optimized for faster transactions (e.g. inserts, updates & deletes)
  - Tables normalized to reduce/eliminate redundancies
  - Table schemas can be hard for business users to understand

Data Modeling using a Star Schema

- OLAP Modeling often based on Star Schema
  - Tables defined as fact tables or dimension tables
  - Fact tables related to dimension table using 1-to-many relationships

Designing Queries to Build a Star Schema

- Converts OLTP Data Model to OLAP Data Model
  - Sales table is modeled as a OLAP Fact Table
  - Other tables are modeled as OLAP Dimension tables
  - Requires pulling CustomerId column into Sales table
  - All dimension tables should be directly related to fact table

DEMO

Exploring the Wingtip Sales Analysis Demo Project

Agenda

- ✓ Deciding What To Measure
- ✓ Query Design Fundamentals
- ✓ Designing Data Model using a Star Schema
- ➤ Working with the Query Editor Window
- • Importing Content From SharePoint Online
- • Designing with Function Queries
- • Understanding Parameters and Template Files

Query Editor Ribbon Tabs

Home tab

Transform tab

Add Column tab

View tab

Examples of Basic Power BI Desktop Steps

- Rename column
- Convert column type
- Trim and clean column values
- Replace column values
- Format column values
- Expanding related column
- Merging columns
- Splitting columns

Cleaning Data

- Special steps available to clean up string-based data
  - **Transform > Trim** removes whitespace
  - **Transform > Clean** removed non-printable characters

Converting Column Types

- Transform data to make it more reliable
  - Convert date-time column to date column
- Transform data to make it more efficient
  - Convert decimal to fixed decimal number for currency

PurchaseDate	1²3 Quantity	$ SalesAmount	$ ProductCost	
1/28/2012	1	2.95	1.2 Decimal Number
1/28/2012	6		$ Fixed Decimal Number
1/28/2012	1	19.95	1²3 Whole Number
1/28/2012	5	249.75	Date/Time
1/28/2012	1	2.95	Date

Expanding Related Columns

- Used to pull data from related tables
    - Saves you from performing SQL joins or VLOOKUP

Adding a Custom Column

- Custom column provide custom logic
  - Logic must be written in M programming language

Adding a Conditional Column

- Abstracts away need to write M code

DEMO

Using Queries to Transform Data During the Load Process

Agenda

- ✓ Deciding What To Measure
- ✓ Query Design Fundamentals
- ✓ Designing Data Model using a Star Schema
- ✓ Working with the Query Editor Window
- ➢ Importing Content From SharePoint Online
- • Designing with Function Queries
- • Understanding Parameters and Template Files

Importing Files using the Web Datasource

- Files in SharePoint document library exposed via HTTPS
  - Use **Web** datasource to import files in SharePoint Online
  - Use the absolute path to file in document library
  - Authenticate using **Organizational account**
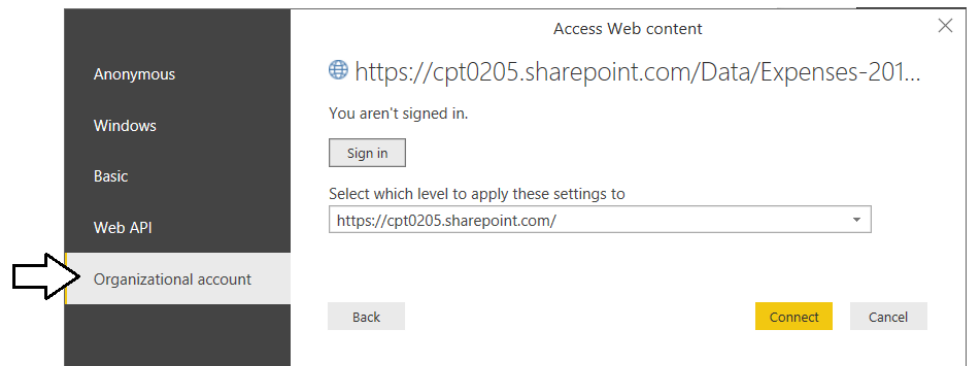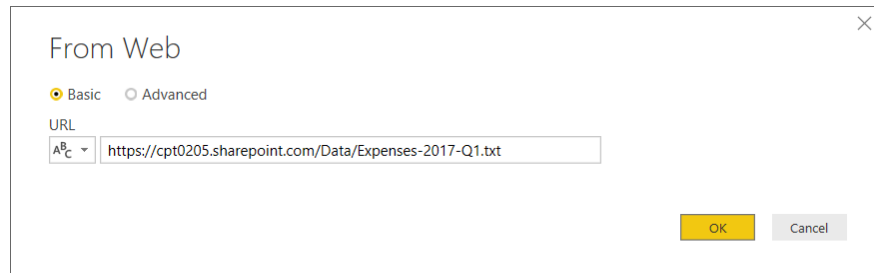
Importing using the SharePoint Folder

- Select the **SharePoint folder** datasource

- Query returns a row for each file in the site

DEMO

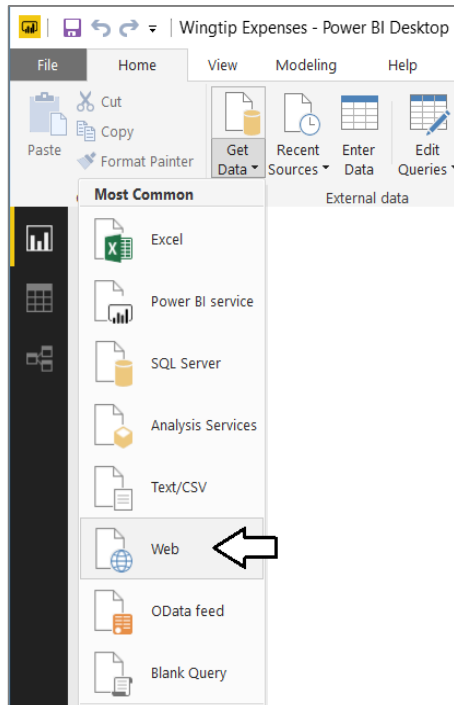**Importing Content from Files in SharePoint Online**

Agenda

- ✓ Deciding What To Measure
- ✓ Query Design Fundamentals
- ✓ Designing Data Model using a Star Schema
- ✓ Working with the Query Editor Window
- ✓ Importing Content From SharePoint Online
- ➢ Designing with Function Queries
- • Understanding Parameters and Template Files

DEMO

Creating a Function Query

Agenda

- ✓ Deciding What To Measure
- ✓ Query Design Fundamentals
- ✓ Designing Data Model using a Star Schema
- ✓ Working with the Query Editor Window
- ✓ Importing Content From SharePoint Online
- ✓ Designing with Function Queries
- ➢ Understanding Parameters and Template Files

Query Parameters

- What is a Query Parameter?
    - Configurable setting with project scope
    - Strongly-typed value to which you can apply restrictions
    - Can be referenced from a query
    - Can be referenced from DAX code in data model

- Where are Parameters commonly used
    - To parameterize data source connection details
    - To filter rows when importing data

Creating Query Parameters

- Parameters can be created using **Manager Parameters** menu
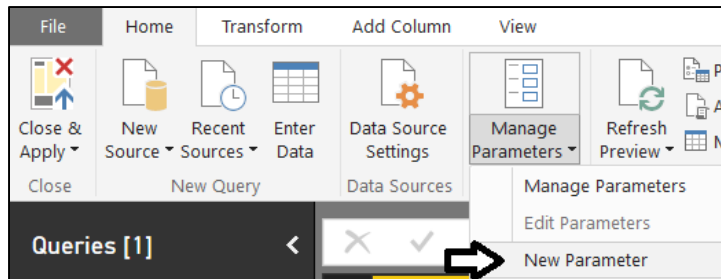
- Parameter properties
  - Name
  - Description
  - Required
  - Allowed Values
  - Default Value
  - Current Value

Referencing Parameters in a Query

- Parameters can be referenced inside query
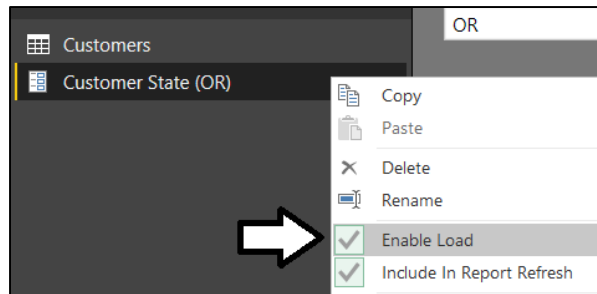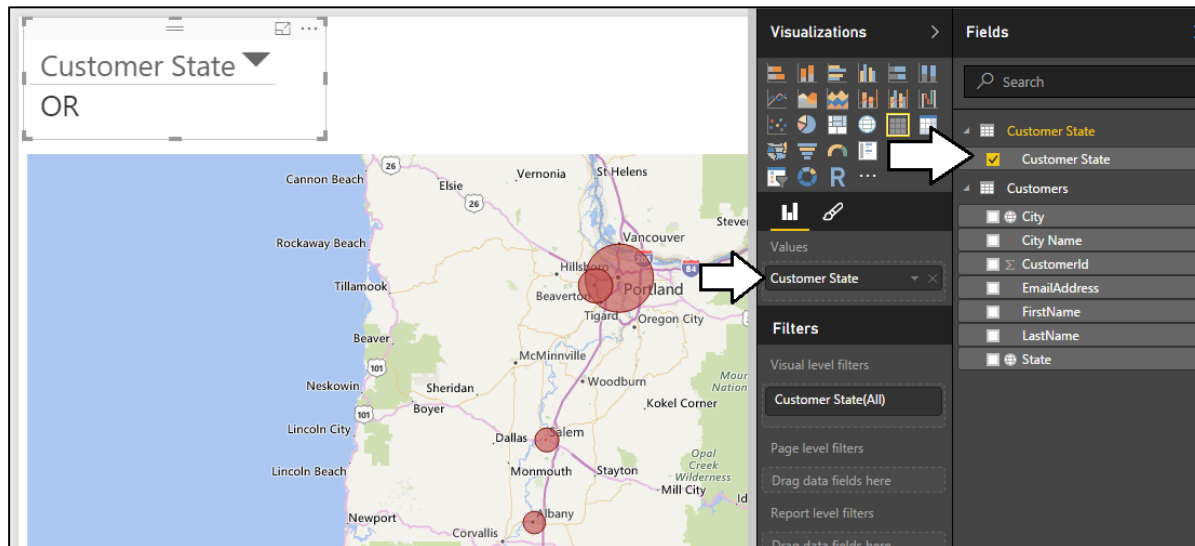 - Next query execution uses current parameter value

Making Parameters Available to Data Model

- Configure parameter's Enable Load setting

- Parameter becomes visible within fields list in report view

Power BI Project Template Files

- PBIX project can be exported to project template file
  - Template file created with PBIT file extension
  - Generated template files contains everything except for the data
  - PBIT template file can be imported to create new PBIX projects
  - Template files are powerful when used together with parameters

- How are template files used?
  - Export PBIX project to create a PBIT template file
  - Import the PBIT template file to create a new PBIX project
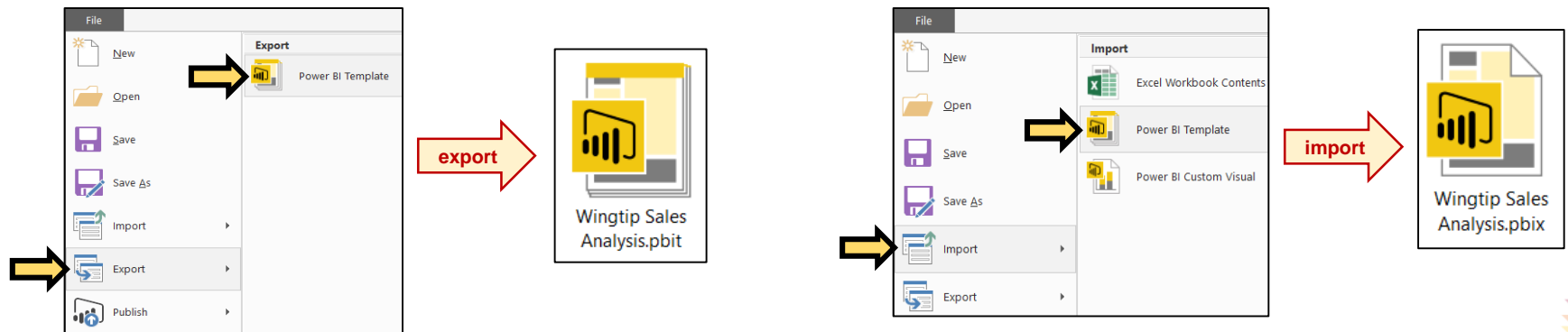
Understanding Function Queries

- Query can be converted into reusable function
  - Requires editing query M code in Advanced Editor
  - Function query can be defined to accept parameters


```
GetExpensesFromFile

(FilePath as text) =>

let
    Source = Csv.Document(Web.Contents(FilePath)
    #"Changed Type" = Table.TransformColumnTypes
```

  - Function query can't be edited with visual designer

Summary

- ✓ Deciding What To Measure
- ✓ Query Design Fundamentals
- ✓ Designing Data Model using a Star Schema
- ✓ Working with the Query Editor Window
- ✓ Importing Content From SharePoint Online
- ✓ Designing with Function Queries
- ✓ Understanding Parameters and Template Files