

POWER BI TRAINING



Power BI

TURN YOUR DATA INTO IMPACT!

HEALTHCARE CASE STUDY - (FOR TRANSFORMATION)

THE SITUATION

You are hired by the CEO of a business consulting company who are trying to analyse the healthcare industry. They want you to deep dive into their reports and create a comprehensive Business Intelligence (BI) solution. As a data analyst you are given access to the reports.

THE BRIEF

Your CEO wants you to dig deep into the data and bring out interesting talking points, compare regional performance, analyse product level trends and forecasts based on the data.

THE OBJECTIVE

In this section use Power BI desktop to -

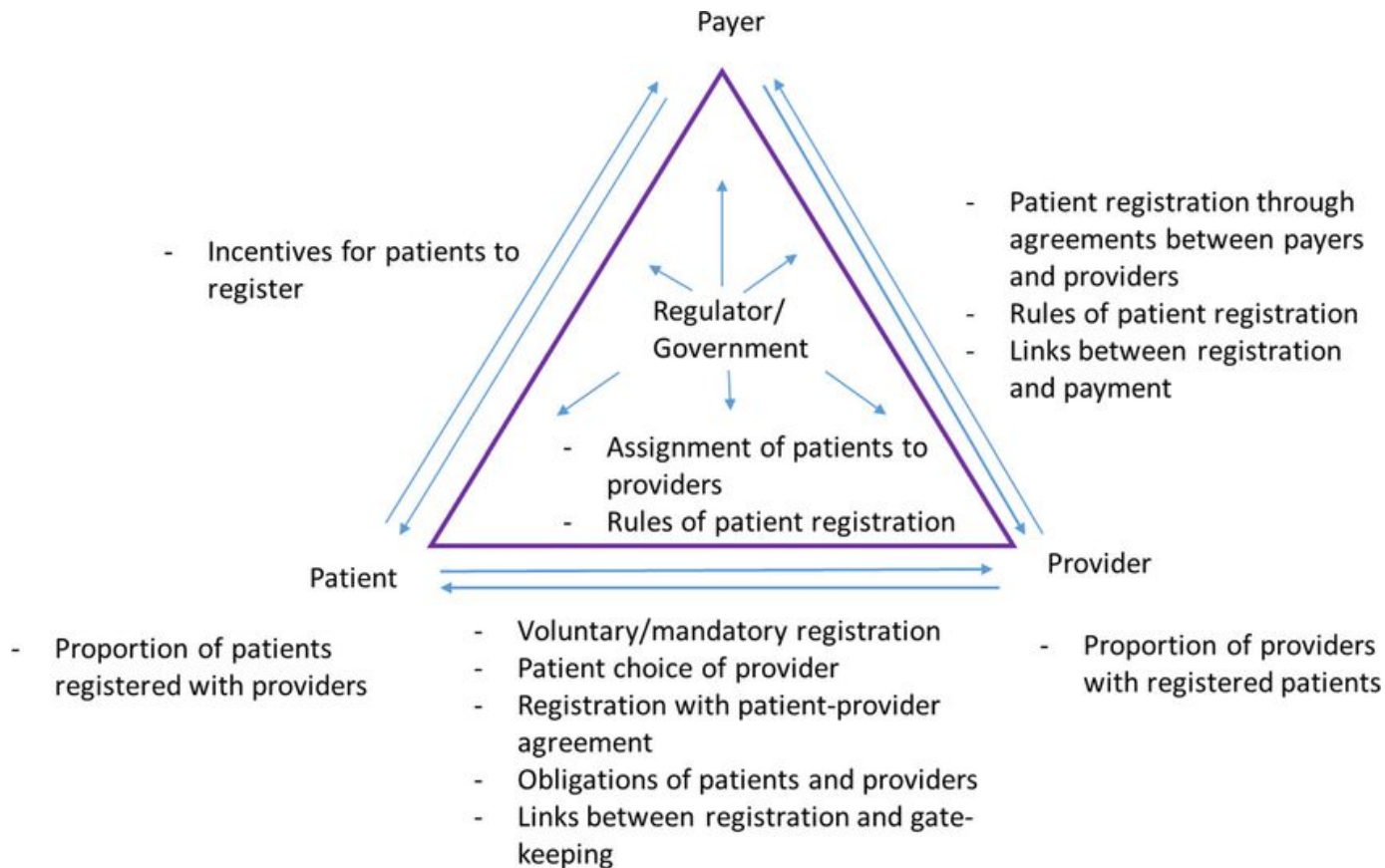
- Connect the Healthcare dataset
- Transform the raw data

ABOUT THE HEALTHCARE DATASET

The given dataset is a collection of data related to patients, healthcare providers, medical procedures, insurance, and payments for the US Market. It encompasses details such as patient demographics (such as names, gender, age, country, city, and state), provider information (including names and specialties), dates of medical procedure postings, descriptions of medical procedures and their groupings, insurance payer names, hospital names, diagnosis code descriptions and groupings, quantities of medical procedures performed, expenses, adjustments, insurance payments, and patient payments.

This dataset serves as a valuable resource for analyzing and understanding various aspects of healthcare data. By exploring this dataset, one can gain a deeper understanding of the healthcare industry and how does analytics help in valuable insights for the stakeholders. It further helps the learner to understand how the interactions between patients, providers, insurance companies, and healthcare systems lead to informed decision-making and improvements in healthcare delivery.

HEALTHCARE INDUSTRY PLAYERS



DATA DICTIONARY - DATA ABOUT DATA

A Data Dictionary is a collection of names, definitions, and attributes about data elements that are being used or captured in a database. It contains metadata i.e. data about the database.

Data dictionary usually contains -

- Names of all the database tables and their schemas.
- Details about all the tables in the database, such as their owners, their security constraints, when they were created etc.
- Physical information about the tables such as where they are stored and how.
- Table constraints such as primary key attributes, foreign key information etc.
- Information about the database views that are visible.

ABOUT THE DATASET - DATA DICTIONARY

| Field Name | Description |
|--------------------------|---|
| FactTableID | An identifier for each record in the dataset. |
| PatientID | Unique identifier for each patient. |
| patientFirstName | First name of the patient. |
| patientLastName | Last name of the patient. |
| patientEmail | Email address of the patient. |
| PatientGender | Gender of the patient. |
| PatientAge | Age of the patient. |
| patientCountry | Country where the patient is located. |
| patientCity | City where the patient is located. |
| patientState | State where the patient is located. |
| ProviderName | Name of the healthcare provider. |
| ProviderSpecialty | Specialty or field of expertise of the healthcare provider. |

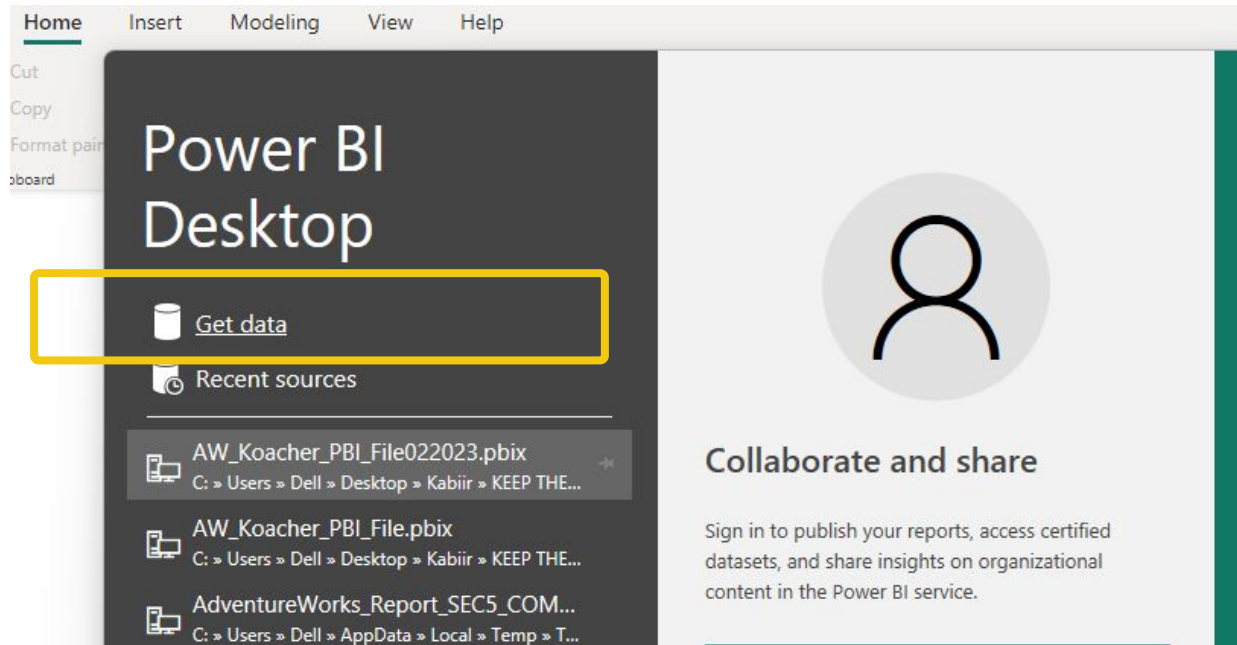
ABOUT THE DATASET - DATA DICTIONARY

| Field Name | Description |
|---------------------------------|---|
| Date of posting | Date when the medical procedure or service was posted. |
| CptDesc | Description of the Current Procedural Terminology (CPT) code for the medical procedure. |
| CptGrouping | Grouping or category of the CPT code. |
| PayerName | Name of the insurance payer or provider. |
| Hospital name | Name of the hospital where the medical procedure took place. |
| DiagnosisCodeDescription | Description of the diagnosis code related to the medical procedure. |
| DiagnosisCodeGroup | Group or category of the diagnosis code. |
| CPTUnits | Units or quantity of the medical procedure performed. |
| Expenses | Expenses or costs associated with the medical procedure. |
| Adjustment | Adjustment made to the expenses or costs. |
| Insurance_Payment | Payment received from the insurance provider. |
| Patient_Payment | Payment made by the patient. |

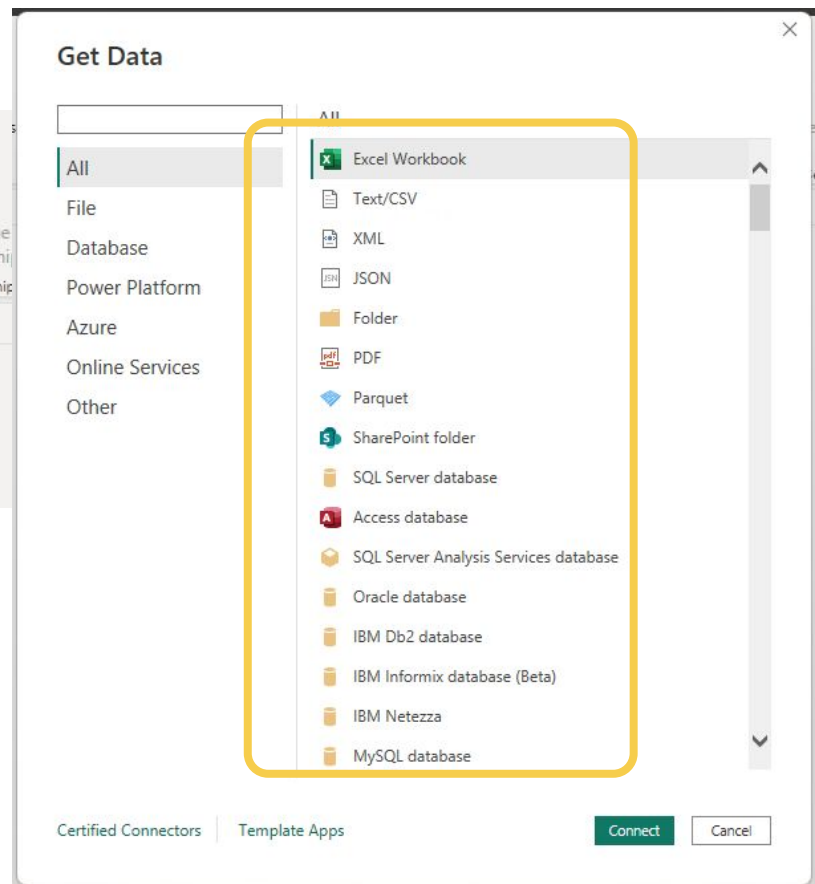
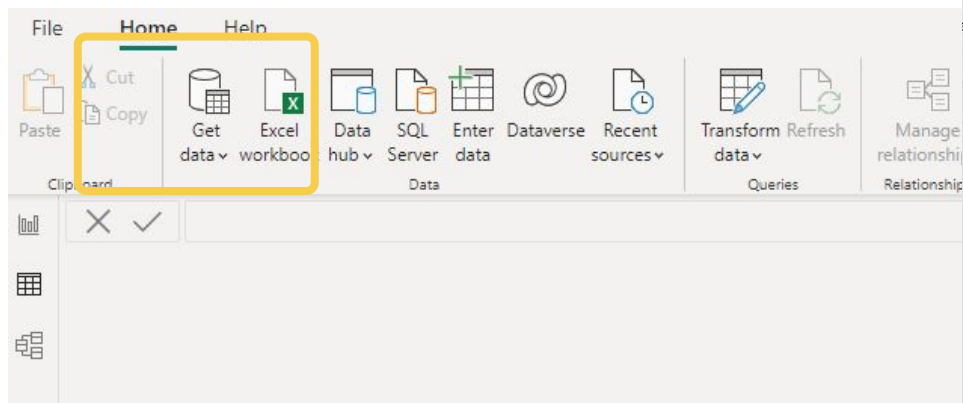
CONNECTING AND RESHAPING DATA

HANDS ON - CONNECTING DATA

1. Connect PowerBI to '[health_care_dataset.csv](#)' file



CONNECT FROM HOME TAB



HANDS ON - QUERY EDITOR

Health_care_dataset.csv

File Origin: 1252: Western European (Windows) | Delimiter: Comma | Data Type Detection: Based on first 200 rows

| FactTableID | PatientID | patientFirstName | patientLastName | patientEmail | PatientGender | PatientAge | patientCountry |
|-------------|-----------|------------------|-----------------|---------------------------------|---------------|------------|----------------|
| 1082164042 | 6066415 | Jacob | Sanderson | jacob.sanderson@datacourse.com | Male | 45 | United States |
| 1027823678 | 5810482 | Jason | Rees | jason.rees@datacourse.com | Male | 20 | United States |
| 922963624 | 5337135 | Ruth | Campbell | ruth.campbell@datacourse.com | Female | 65 | United States |
| 923120306 | 5337135 | Ruth | Campbell | ruth.campbell@datacourse.com | Female | 65 | United States |
| 951760836 | 5337135 | Ruth | Campbell | ruth.campbell@datacourse.com | Female | 65 | United States |
| 939266959 | 5456888 | Anna | McDonald | anna.mcdonald@datacourse.com | Female | 29 | United States |
| 1082333983 | 6066415 | Jacob | Sanderson | jacob.sanderson@datacourse.com | Male | 45 | United States |
| 1078743980 | 5251861 | Brian | Ellison | brian.ellison@datacourse.com | Male | 40 | United States |
| 978466670 | 5251861 | Brian | Ellison | brian.ellison@datacourse.com | Male | 40 | United States |
| 977806621 | 5723953 | Anna | Jackson | anna.jackson@datacourse.com | Female | 69 | United States |
| 1004185717 | 5475922 | Zoe | Martin | zoe.martin@datacourse.com | Female | 1 | United States |
| 980580337 | 5343970 | Benjamin | Russell | benjamin.russell@datacourse.com | Male | 82 | United States |
| 981056255 | 5343970 | Benjamin | Russell | benjamin.russell@datacourse.com | Male | 82 | United States |
| 1068435771 | 5952218 | Olivia | Quinn | olivia.quinn@datacourse.com | Female | 39 | United States |
| 1078969155 | 5952218 | Olivia | Quinn | olivia.quinn@datacourse.com | Female | 39 | United States |
| 1093656816 | 5952218 | Olivia | Quinn | olivia.quinn@datacourse.com | Female | 39 | United States |
| 1119072252 | 5814278 | Amanda | Blake | amanda.blake@datacourse.com | Female | 69 | United States |
| 1124233962 | 5814278 | Amanda | Blake | amanda.blake@datacourse.com | Female | 69 | United States |
| 927135438 | 5502929 | Liam | Cameron | liam.cameron@datacourse.com | Male | 53 | United States |
| 92939787 | 5502929 | Liam | Cameron | liam.cameron@datacourse.com | Male | 53 | United States |

Extract Table Using Examples | Load | Transform Data | Cancel

- Select the appropriate Data Type detection.
- Click on the 'Transform Data' to open the Query Editor window.

(Clicking on the 'Load' button will load the data into the main interface where you view, create models and build visualizations. But as of now we will proceed to the Power Query Editor.)

THE QUERY EDITOR



Query Editing Tools (Table transformations, calculated columns, etc)

Formula Bar
(this is "M" code)

**Query
Pane**

AdventureWorks_Report_COMPLETE - Power Query Editor

File Home Transform Add Column View Tools Help

Close & Apply * New Source * Recent Sources * Enter Data * Data source settings * Manage Parameters * Refresh Preview * Advanced Editor * Choose Columns * Remove Columns * Manage Columns * Keep Rows * Remove Rows * Reduce Rows * Sort * Split Column * Group By * Data Type: Whole Number * Use First Row as Headers * Replace Values * Merge Queries * Append Queries * Combine Files * Combine

Formula Bar: = Table.RemoveColumns(#\"Filtered Rows\", {\"BirthYear\"})

Query Pane: Transform File from AW_Sales [3] Other Queries [8] AW_Product_Lookup AW_Customer_Lookup AW_Calendar_Lookup AW_Sales AW_Territories_Lookup AW_Product_Category_Lookup AW_Product_Subcategory_Lookup AW>Returns

| CustomerKey | Prefix | FirstName | LastName | BirthDate |
|-------------|------------|-----------|----------|------------|
| 1 | 11000 Mr. | Jan | Yang | 4/8/1965 |
| 2 | 11001 Mr. | Eugene | Huang | 5/14/1965 |
| 3 | 11002 Mr. | Ruben | Tomas | 8/12/1965 |
| 4 | 11003 Mr. | Christy | Zhu | 2/15/1968 |
| 5 | 11004 Mrs. | Elizabeth | Johnson | 8/6/1968 |
| 6 | 11005 Mr. | Julio | Ruiz | 8/9/1969 |
| 7 | 11007 Mr. | Marco | Melita | 5/9/1964 |
| 8 | 11008 Mrs. | Rubin | Vierhoff | 7/7/1964 |
| 9 | 11009 Mr. | Shannon | Carlson | 4/1/1964 |
| 10 | 11010 Mr. | Jacquelyn | Suarez | 2/6/1964 |
| 11 | 11011 Mr. | Curtis | Lu | 11/4/1963 |
| 12 | 11012 Mrs. | Lauren | Walker | 1/18/1968 |
| 13 | 11013 Mr. | Ian | Jenkins | 8/6/1968 |
| 14 | 11014 Mrs. | Sydney | Bennett | 5/9/1968 |
| 15 | 11015 Mr. | Chloe | Young | 2/27/1979 |
| 16 | 11016 Mr. | Wyllis | Hill | 4/26/1979 |
| 17 | 11017 Mrs. | Shannon | Wang | 6/26/1944 |
| 18 | 11018 Mr. | Clarence | Raji | 10/9/1944 |
| 19 | 11019 Mr. | Luke | Lai | 1/7/1978 |
| 20 | 11020 Mr. | Jordan | King | 8/20/1978 |
| 21 | 11021 Mr. | Destiny | Wilson | 9/3/1978 |
| 22 | 11022 Mr. | Stuart | Zhang | 10/12/1978 |

Query Settings: Name: AW_Customer_Lookup, All Properties

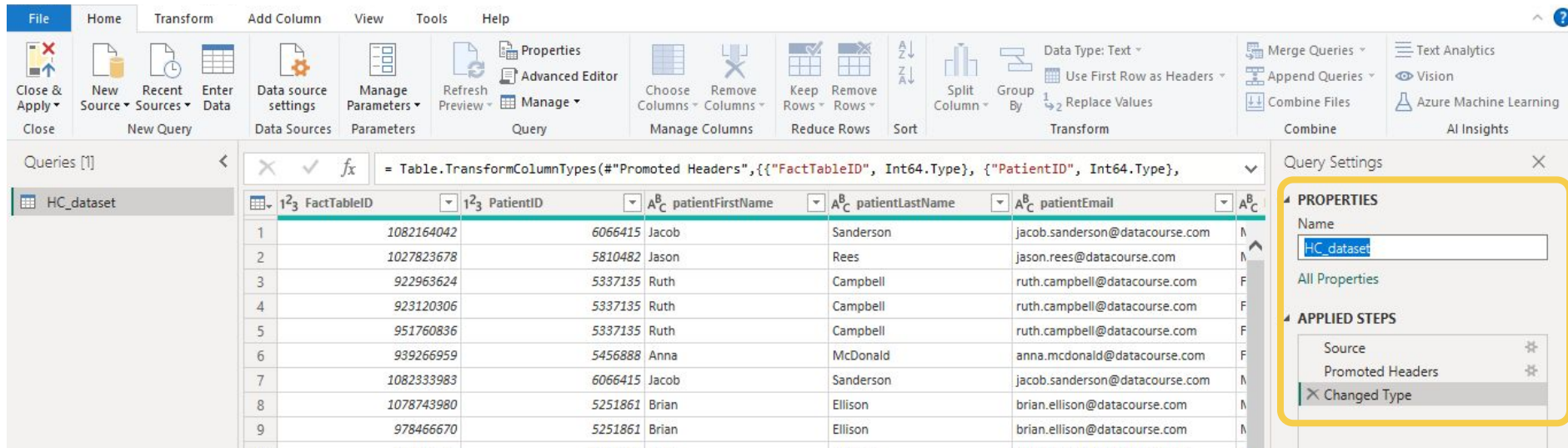
APPLIED STEPS: Source, Promoted Headers, Changed Type, Capitalized Each Word, Inserted FullName Column, Inserted Text Before Delimiter, Renamed Columns, Inserted Text Between Delimit..., Renamed Columns1, Replaced Value, Capitalized Each Word1, Inserted Year, Renamed Columns2, Added Conditional Column, Filtered Rows, Removed Columns



Power BI

HANDS ON - QUERY EDITOR

1. Change the name of the file to - [HC_dataset.csv](#).
2. Browse through all the data types and see if they are good.
3. Also verify the Applied steps automatically performed by PBI



The screenshot displays the Power BI Query Editor interface. The top ribbon includes tabs for File, Home, Transform, Add Column, View, Tools, and Help. The 'Transform' tab is active, showing various data transformation options like 'Choose Columns', 'Remove Columns', 'Keep Rows', 'Remove Rows', 'Sort', 'Split Column', 'Group By', 'Data Type: Text', 'Use First Row as Headers', 'Replace Values', 'Merge Queries', 'Append Queries', 'Combine Files', 'Text Analytics', 'Vision', and 'Azure Machine Learning'.

The main area shows a table with the following columns: FactTableID, PatientID, patientFirstName, patientLastName, and patientEmail. The data is filtered to show 9 rows. The 'Query Settings' pane on the right is open, showing the 'PROPERTIES' section with the name 'HC_dataset' and the 'APPLIED STEPS' section listing 'Source', 'Promoted Headers', and 'Changed Type'.

| FactTableID | PatientID | patientFirstName | patientLastName | patientEmail |
|-------------|------------|------------------|-----------------|--------------------------------|
| 1 | 1082164042 | Jacob | Sanderson | jacob.sanderson@datacourse.com |
| 2 | 1027823678 | Jason | Rees | jason.rees@datacourse.com |
| 3 | 922963624 | Ruth | Campbell | ruth.campbell@datacourse.com |
| 4 | 923120306 | Ruth | Campbell | ruth.campbell@datacourse.com |
| 5 | 951760836 | Ruth | Campbell | ruth.campbell@datacourse.com |
| 6 | 939266959 | Anna | McDonald | anna.mcdonald@datacourse.com |
| 7 | 1082333983 | Jacob | Sanderson | jacob.sanderson@datacourse.com |
| 8 | 1078743980 | Brian | Ellison | brian.ellison@datacourse.com |
| 9 | 978466670 | Brian | Ellison | brian.ellison@datacourse.com |

DATA TABLES AND LOOK-UP TABLES

Data Tables (also called Fact tables) - Record business transactions which contains numbers and values. They are granular tables with 'ID' or 'Key' fields (also known as identifiers). These identifiers are used to create relationship between these tables and 'Look up Tables'. Example of Data tables - Sales table, Returns table, Bank Loan transactions, Patients treatment records, Subscription bought etc.

Look Up Tables contains the descriptions and details of categorical variables in the business transactions. These are usually text based attributes about Dimensions in the data e.g - Customer, Products, Patients, Region etc.

| date | product_id | quantity |
|----------|------------|----------|
| 1/1/1997 | 869 | 5 |
| 1/1/1997 | 1472 | 3 |
| 1/1/1997 | 76 | 4 |
| 1/1/1997 | 320 | 3 |
| 1/1/1997 | 4 | 4 |
| 1/1/1997 | 952 | 4 |
| 1/1/1997 | 1222 | 4 |
| 1/1/1997 | 517 | 4 |
| 1/1/1997 | 1359 | 4 |
| 1/1/1997 | 357 | 4 |
| 1/1/1997 | 1426 | 5 |
| 1/1/1997 | 190 | 4 |
| 1/1/1997 | 367 | 4 |
| 1/1/1997 | 250 | 5 |
| 1/1/1997 | 600 | 4 |
| 1/1/1997 | 702 | 5 |

This Data Table contains "quantity" values, and connects to lookup tables via the "date" and "product_id" columns

| date | day_of_month | month | year | weekday | week_of_year | week_ending | month_name | quarter |
|----------|--------------|-------|------|-----------|--------------|-------------|------------|---------|
| 1/1/1997 | 1 | 1 | 1997 | Wednesday | 1 | 1/5/1997 | January | Q1 |
| 1/2/1997 | 2 | 1 | 1997 | Thursday | 1 | 1/5/1997 | January | Q1 |
| 1/3/1997 | 3 | 1 | 1997 | Friday | 1 | 1/5/1997 | January | Q1 |
| 1/4/1997 | 4 | 1 | 1997 | Saturday | 1 | 1/5/1997 | January | Q1 |
| 1/5/1997 | 5 | 1 | 1997 | Sunday | 2 | 1/5/1997 | January | Q1 |
| 1/6/1997 | 6 | 1 | 1997 | Monday | 2 | 1/12/1997 | January | Q1 |

This Calendar Lookup table provides additional attributes about each date (month, year, weekday, quarter, etc.)

| product_id | product_brand | product_name | product_sku | product_retail_price | product_cost | product_weight |
|------------|---------------|-----------------------------|-------------|----------------------|--------------|----------------|
| 1 | Washington | Washington Berry Juice | 90748583674 | 2.85 | 0.94 | 8.39 |
| 2 | Washington | Washington Mango Drink | 96516502499 | 0.74 | 0.26 | 7.42 |
| 3 | Washington | Washington Strawberry Drink | 58427771925 | 0.83 | 0.4 | 13.1 |
| 4 | Washington | Washington Cream Soda | 64412155747 | 3.64 | 1.64 | 10.6 |
| 5 | Washington | Washington Diet Soda | 85561191439 | 2.19 | 0.77 | 6.66 |
| 6 | Washington | Washington Cola | 29804642796 | 1.15 | 0.37 | 15.8 |
| 7 | Washington | Washington Diet Cola | 20191444754 | 2.61 | 0.91 | 18 |
| 8 | Washington | Washington Orange Juice | 89770532250 | 2.59 | 0.8 | 8.97 |

This Product Lookup table provides additional attributes about each product (brand, product name, sku, price, etc.)

PRIMARY KEYS AND FOREIGN KEYS

| date | product_id | quantity |
|----------|------------|----------|
| 1/1/1997 | 869 | 5 |
| 1/1/1997 | 1472 | 3 |
| 1/1/1997 | 76 | 4 |
| 1/1/1997 | 320 | 3 |
| 1/1/1997 | 4 | 4 |
| 1/1/1997 | 952 | 4 |
| 1/1/1997 | 1222 | 4 |
| 1/1/1997 | 517 | 4 |
| 1/1/1997 | 1359 | 4 |
| 1/1/1997 | 357 | 4 |
| 1/1/1997 | 1426 | 5 |
| 1/1/1997 | 190 | 4 |
| 1/1/1997 | 367 | 4 |
| 1/1/1997 | 250 | 5 |
| 1/1/1997 | 600 | 4 |
| 1/1/1997 | 702 | 5 |

| date | day_of_month | month | year | weekday | week_of_year | week_ending | month_name | quarter |
|----------|--------------|-------|------|-----------|--------------|-------------|------------|---------|
| 1/1/1997 | 1 | 1 | 1997 | Wednesday | 1 | 1/5/1997 | January | Q1 |
| 1/2/1997 | 2 | 1 | 1997 | Thursday | 1 | 1/5/1997 | January | Q1 |
| 1/3/1997 | 3 | 1 | 1997 | Friday | 1 | 1/5/1997 | January | Q1 |
| 1/4/1997 | 4 | 1 | 1997 | Saturday | 1 | 1/5/1997 | January | Q1 |
| 1/5/1997 | 5 | 1 | 1997 | Sunday | 2 | 1/5/1997 | January | Q1 |
| 1/6/1997 | 6 | 1 | 1997 | Monday | 2 | 1/12/1997 | January | Q1 |

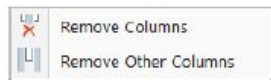
| product_id | product_name | product_sku | product_retail_price | product_cost | product_weight |
|------------|-----------------------------|-------------|----------------------|--------------|----------------|
| 1 | Washington Berry Juice | 90748583674 | 2.85 | 0.94 | 8.39 |
| 2 | Washington Mango Drink | 96516502499 | 0.74 | 0.26 | 7.42 |
| 3 | Washington Strawberry Drink | 58427771925 | 0.83 | 0.4 | 13.1 |
| 4 | Washington Cream Soda | 64412155747 | 3.64 | 1.64 | 10.6 |
| 5 | Washington Diet Soda | 85561191439 | 2.19 | 0.77 | 6.66 |
| 6 | Washington Cola | 29804642796 | 1.15 | 0.37 | 15.8 |
| 7 | Washington Diet Cola | 20191444754 | 2.61 | 0.91 | 18 |
| 8 | Washington Orange Juice | 89770532250 | 2.59 | 0.8 | 8.97 |

FOREIGN KEYS exist usually in Data/Fact tables*. They contain multiple instances of the identifier keys and are mapped with the Primary keys from the 'Look up Tables'.

PRIMARY KEYS exist usually in 'Lookup tables'*. They are Unique identifiers which are used with Foreign Keys to map the attributes/details of the dimension and create logical relationships between tables

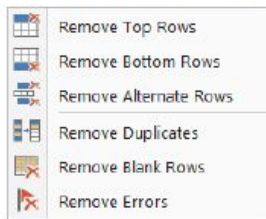
* Foreign and Primary keys can coexist in Data tables also.

QUERY EDITOR - BASIC TRANSFORMATIONS



Choose or remove columns

Tip: use the "Remove Other Columns" option if you always want a specific set

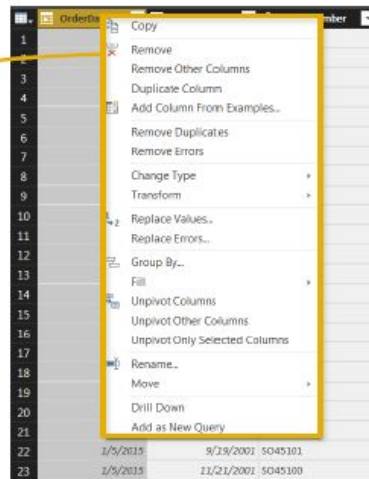


Keep or remove rows

Tip: use the "Remove Duplicates" option to create a new lookup table from scratch

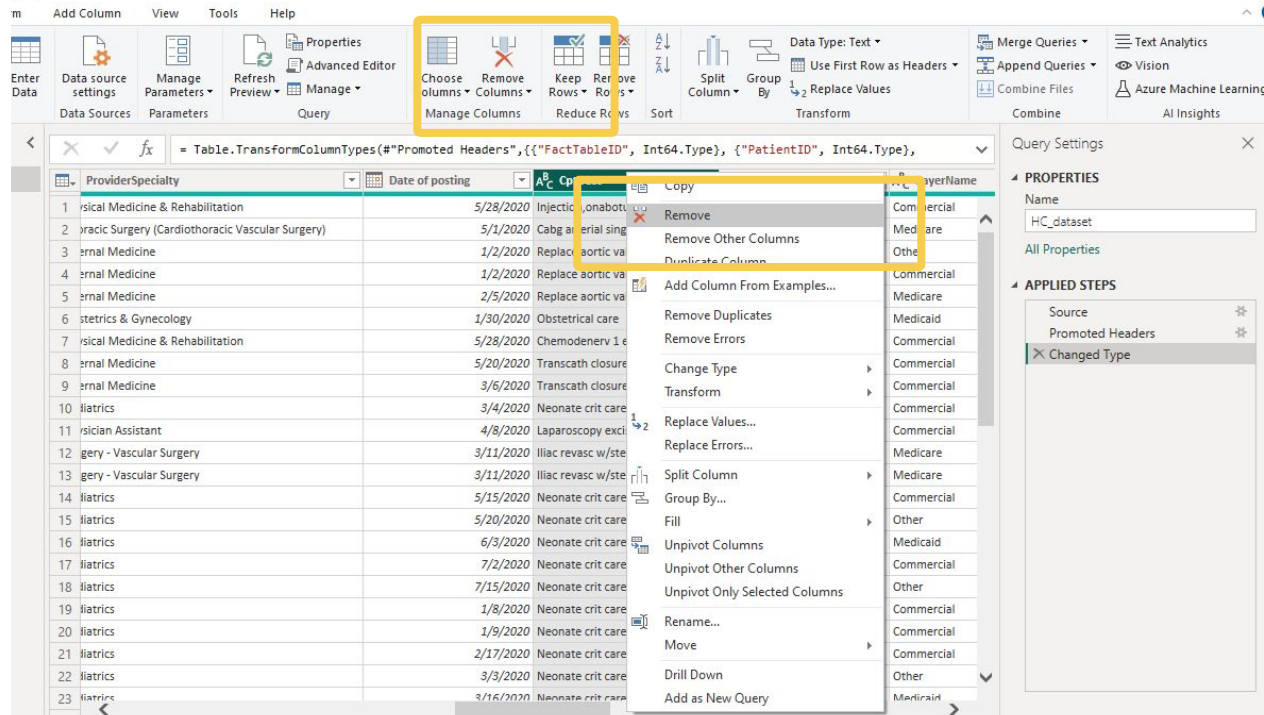
Duplicate, move & rename columns

Tip: Right-click the column header to access common tools



HANDS ON - QUERY EDITOR

Remove the 'CptDesc' column. You can click on the column or use Remove columns in the Menu bar.



The screenshot displays the Power BI Query Editor interface. The top menu bar includes 'Add Column', 'View', 'Tools', and 'Help'. Below this, the 'Query' tab is active, showing a list of columns: 'ProviderSpecialty', 'Date of posting', 'CptDesc', and 'PayerName'. The 'CptDesc' column is highlighted with a yellow box. A context menu is open over the 'CptDesc' column, with the 'Remove' option highlighted. The 'Remove' option is also highlighted in the 'Query' tab's menu bar. The 'Remove' option is highlighted in the 'Query' tab's menu bar. The 'Remove' option is highlighted in the 'Query' tab's menu bar.

| ProviderSpecialty | Date of posting | CptDesc | PayerName |
|--|-----------------|-------------------------|------------|
| Physical Medicine & Rehabilitation | 5/28/2020 | Injection, onabotulinum | Commercial |
| Thoracic Surgery (Cardiothoracic Vascular Surgery) | 5/1/2020 | Cabg aortic valve sing | Medicare |
| Internal Medicine | 1/2/2020 | Replace aortic valve | Other |
| Internal Medicine | 1/2/2020 | Replace aortic valve | Commercial |
| Internal Medicine | 2/5/2020 | Replace aortic valve | Medicare |
| Obstetrics & Gynecology | 1/30/2020 | Obstetrical care | Medicaid |
| Physical Medicine & Rehabilitation | 5/28/2020 | Chemodendromy | Commercial |
| Internal Medicine | 5/20/2020 | Transcath closure | Commercial |
| Internal Medicine | 3/6/2020 | Transcath closure | Commercial |
| Neonatology | 3/4/2020 | Neonate crit care | Commercial |
| Physician Assistant | 4/8/2020 | Laparoscopy exci | Commercial |
| Surgery - Vascular Surgery | 3/11/2020 | Iliac revasc w/ste | Medicare |
| Surgery - Vascular Surgery | 3/11/2020 | Iliac revasc w/ste | Medicare |
| Neonatology | 5/15/2020 | Neonate crit care | Commercial |
| Neonatology | 5/20/2020 | Neonate crit care | Other |
| Neonatology | 6/3/2020 | Neonate crit care | Medicaid |
| Neonatology | 7/2/2020 | Neonate crit care | Commercial |
| Neonatology | 7/15/2020 | Neonate crit care | Other |
| Neonatology | 1/8/2020 | Neonate crit care | Commercial |
| Neonatology | 1/9/2020 | Neonate crit care | Commercial |
| Neonatology | 2/17/2020 | Neonate crit care | Commercial |
| Neonatology | 3/3/2020 | Neonate crit care | Other |
| Neonatology | 3/16/2020 | Neonate crit care | Medicaid |

HANDS ON - QUERY EDITOR

Select the 'Expenses' column. Go to Home View and click ascending sort.

Untitled - Power Query Editor

Home Transform Add Column View Tools Help

New Source Recent Sources Enter Data Data source settings Manage Parameters Refresh Preview Properties Advanced Editor Manage Choose Columns Remove Columns Keep Rows Remove Rows Sort Split Column Group By Data Type: Whole Number Use First Row as Headers Replace Values Merge Queries Append Queries Combine Files Combine

es [1]

fx = Table.Sort(#"Removed Columns",{{"ProductKey", Order.Ascending}})

| | ProductKey | ProductSubcategoryKey | ProductSKU | ProductName | ModelName |
|---|------------|-----------------------|------------|-------------------------|-----------|
| 1 | 214 | 31 | HL-U509-R | Sport-100 Helmet, Red | Sport-100 |
| 2 | 215 | 31 | HL-U509 | Sport-100 Helmet, Black | Sport-100 |

Query Settings

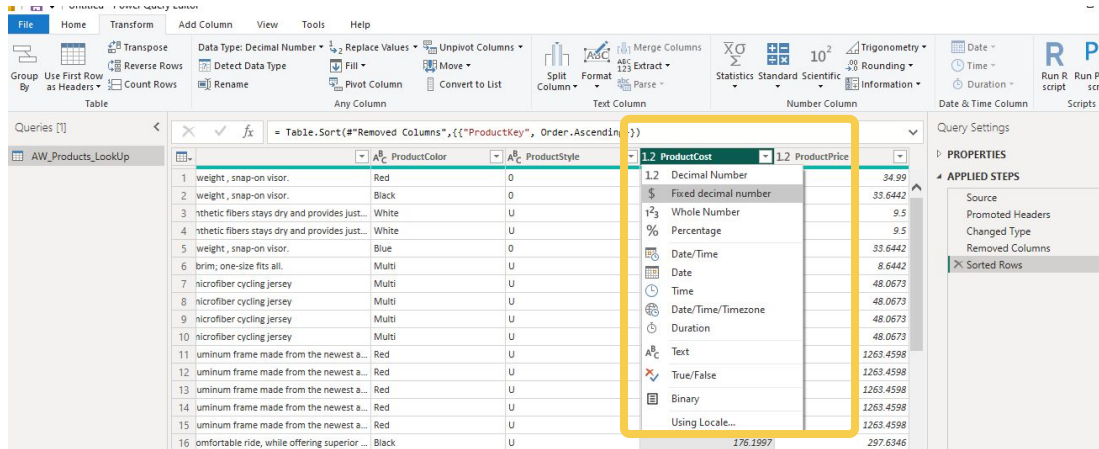
PROPERTIES

APPLIED STEPS

Source

HANDS ON - QUERY EDITOR

1. Left click on the 'Expenses', 'Adjustment', 'Insurance_Payment' and 'Patient_Payment' columns and select data type to - Fixed Decimal Number.
2. Click on the Applied step and change the name of the step as per your requirement.

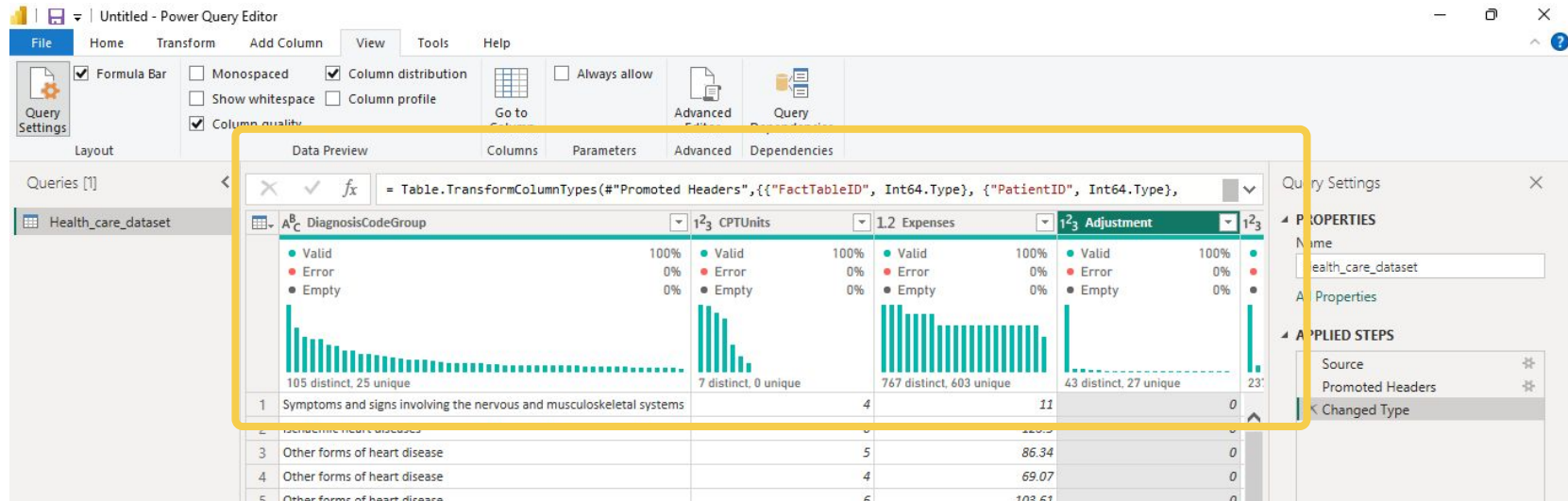


The screenshot shows the Power BI Query Editor interface. The 'Applied Steps' pane on the right is expanded, showing a list of steps: Source, Promoted Headers, Changed Type, Removed Columns, and Sorted Rows. The 'Sorted Rows' step is currently selected and highlighted. The main query area displays a table with columns: ProductColor, ProductStyle, ProductCost, and ProductPrice. The 'ProductCost' column is highlighted, and a context menu is open, showing various data types. The 'Fixed decimal number' option is selected in the menu.

| ProductColor | ProductStyle | ProductCost | ProductPrice |
|--------------|--------------|-------------|--------------|
| Red | 0 | 34.99 | |
| Black | 0 | 33.6442 | |
| White | U | 9.5 | |
| White | U | 9.5 | |
| Blue | 0 | 33.6442 | |
| Multi | U | 8.6442 | |
| Multi | U | 48.0673 | |
| Multi | U | 48.0673 | |
| Multi | U | 48.0673 | |
| Multi | U | 48.0673 | |
| Red | U | 1263.4598 | |
| Red | U | 1263.4598 | |
| Red | U | 1263.4598 | |
| Red | U | 1263.4598 | |
| Red | U | 1263.4598 | |
| Black | U | 297.6346 | |

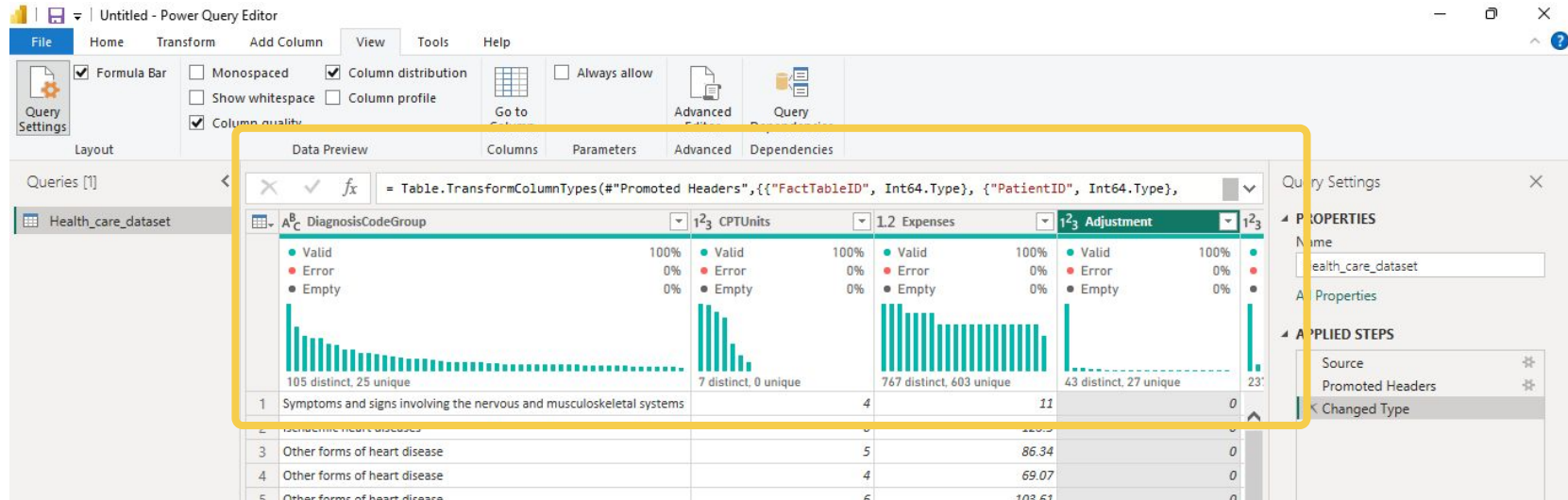
HANDS ON - COLUMN PROFILES

1. Go to View and check for the Column profiles, column distribution and column quality.
2. Analyse the Column profiles for - Expense, Adjustment, Insurance_Payment and Patient_Payment using 1000 rows and entire dataset

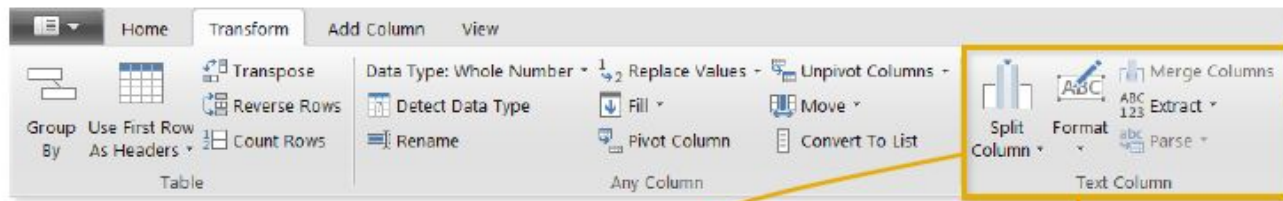


HANDS ON - TREATMENT OF ERRORS

1. Make sure the column profiling is done using entire dataset
2. Now In the 'Adjustment' column, view the rows having errors. Analyse if these are required to be kept or deleted.
3. Keep the errors and replace the Error value by '0'.

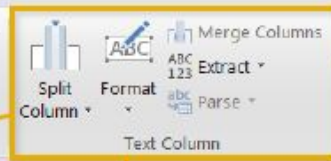


QUERY EDITOR - TEXT TRANSFORMATIONS



| |
|-------------------------|
| By Delimiter |
| By Number of Characters |

Split a text column based on either a specific delimiter or a number of characters



| |
|----------------------|
| lowercase |
| UPPERCASE |
| Capitalize Each Word |
| Trim |
| Clean |
| Add Prefix |
| Add Suffix |

Format a text column to upper, lower or proper case, or add a prefix or suffix

Tip: Use "Trim" to eliminate leading & trailing spaces, or "Clean" to remove non-printable characters

| |
|-------------------------|
| Length |
| First Characters |
| Last Characters |
| Range |
| Text Before Delimiter |
| Text After Delimiter |
| Text Between Delimiters |

Extract characters from a text column based on fixed lengths, first/last, ranges or delimiters

Tip: Select two or more columns to merge (or concatenate) fields

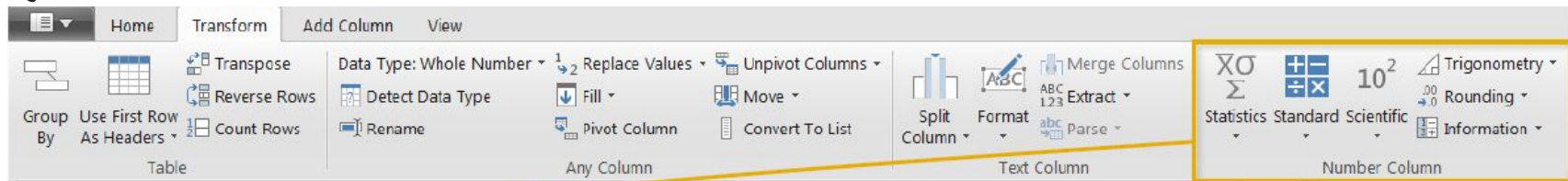
'Transform' and 'Add Column' have many similar tools.
'Transform' tab modifies the existing columns using the tools and 'Add Column' adds a new column by applying the tools.

HANDS ON TEXT TRANSFORMATIONS - QUERY EDITOR

Perform the following text transformations -

1. Extract the patient's email domain from the "patientEmail" column
2. Capitalize the first letter of each word in the "patientFirstName" column
3. Concatenate the patient's first name and last name into a single column
4. Calculate the length of each patient's email address in the "patientEmail" column
5. Replace a specific word (e.g., "Cramp") in the "DiagnosisCodeDescription" column with another word (e.g., "Pain")?
6. Extract the first 3 letters from patientState and create a new variable - 'State Code'
7. Put the code in Capitals
8. Similarly create a Country Code column and put value USA in it

QUERY EDITOR - NUMERIC TRANSFORMATIONS



| |
|-----------------------|
| Sum |
| Minimum |
| Maximum |
| Median |
| Average |
| Standard Deviation |
| Count Values |
| Count Distinct Values |

Statistics functions allow you to evaluate basic stats for the selected column (sum, min/max, average, count, countdistinct, etc)

| |
|----------------|
| Add |
| Multiply |
| Subtract |
| Divide |
| Integer-Divide |
| Modulo |
| Percentage |
| Percent Of |

Standard

| |
|----------------|
| Absolute Value |
| Power |
| Square Root |
| Exponent |
| Logarithm |
| Factorial |

Scientific

| |
|------------|
| Sine |
| Cosine |
| Tangent |
| Arcsine |
| Arccosine |
| Arctangent |

Trigonometry

Standard, Scientific and Trigonometry tools allow you to apply standard operations (addition, multiplication, division, etc.) or more advanced calculations (power, logarithm, sine, tangent, etc) to each value in a column

| |
|---------|
| Is Even |
| Is Odd |
| Sign |

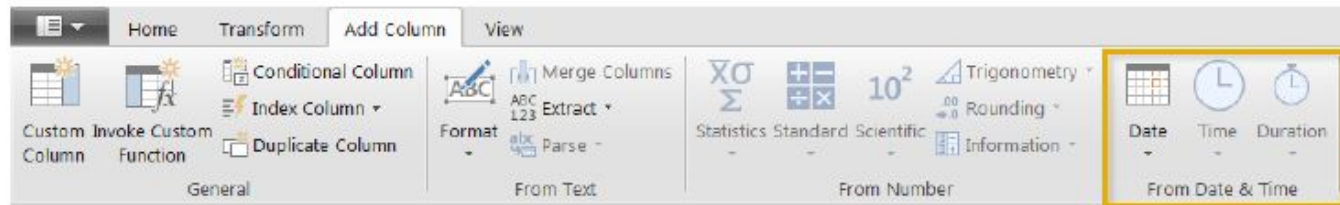
Information tools allow you to define binary flags (TRUE/FALSE or 1/0) to mark each row in a column as even, odd, positive or negative

Note - Stats functions return a SINGLE value and are generally used to explore data while Standard, Scientific and Trigonometry tools apply to each row.

HANDS ON NUMERIC TRANSFORMATIONS - QUERY EDITOR

1. Calculate the average age of the patients.
2. Find the maximum age among the patients.
3. Create a field - 'PatientAge -10'. Subtract 10 from the PatientAge field to create this.
4. Calculate the Standard Deviation of the PatientAge field.
5. Count the number of Male and Female patients.
6. Find the maximum Expenses value in the dataset.
7. Calculate the total number of records in the dataset.
8. Count the total number of cities in the dataset.
9. Create a Column - '% Amount Received'. It is calculated as $(\text{'Insurance_Payment'} + \text{'Patient_Payment'}) / \text{Expenses}$. Apply the unit % to the column.

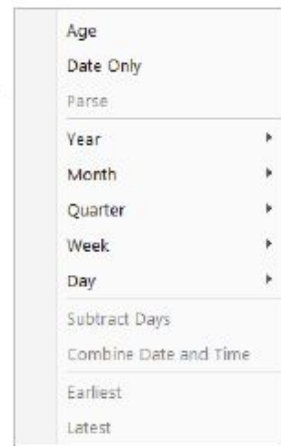
QUERY EDITOR - DATE AND TIME TRANSFORMATIONS



Date & Time tools are relatively straight-forward, and include the following options:

- **Age:** Difference between the current time and the date in each row
- **Date Only:** Removes the time component of a date/time field
- **Year/Month/Quarter/Week/Day:** Extracts individual components from a date field (Time-specific options include Hour, Minute, Second, etc.)
- **Earliest/Latest:** Evaluates the earliest or latest date from a column as a single value (can only be accessed from the "Transform" menu)

Note: You will almost always want to perform these operations from the "Add Column" menu to build out new fields, rather than transforming an individual date/time column



HANDS ON DATE AND TIME TRANSFORMATIONS - QUERY EDITOR

1. Create a new Column "Year".
2. Similarly create 2 more new columns "Month" from the "Date".
3. Create a Column 'Short Month Name' with only 3 letters
4. Create a 'Day' column with the name of the day.
5. Create a Column 'Short Day Name' with only 3 letters
6. Pick out the most busy day of the week in terms of the most patients based on the data?

QUERY EDITOR - INDEX



Index Columns contain a list of sequential values that can be used to identify each unique row in a table (typically starting from 0 or 1)

These columns are often used to create **unique IDs** that can be used to form relationships between tables (more on that later!)

| | Index | OrderDate | StockDate | OrderNumber | ProductKey | CustomerKey |
|----|-------|-----------|------------|-------------|------------|-------------|
| 1 | 1 | 1/1/2015 | 9/21/2001 | SO45080 | 332 | 14657 |
| 2 | 2 | 1/1/2015 | 12/5/2001 | SO45079 | 312 | 29255 |
| 3 | 3 | 1/1/2015 | 10/29/2001 | SO45082 | 350 | 11455 |
| 4 | 4 | 1/1/2015 | 11/16/2001 | SO45081 | 338 | 26782 |
| 5 | 5 | 1/2/2015 | 12/15/2001 | SO45083 | 312 | 14947 |
| 6 | 6 | 1/2/2015 | 10/12/2001 | SO45084 | 310 | 29143 |
| 7 | 7 | 1/2/2015 | 12/18/2001 | SO45086 | 314 | 18747 |
| 8 | 8 | 1/2/2015 | 10/9/2001 | SO45085 | 312 | 18746 |
| 9 | 9 | 1/3/2015 | 10/3/2001 | SO45093 | 312 | 18906 |
| 10 | 10 | 1/3/2015 | 9/29/2001 | SO45090 | 310 | 29170 |
| 11 | 11 | 1/3/2015 | 12/11/2001 | SO45088 | 345 | 11398 |
| 12 | 12 | 1/3/2015 | 10/24/2001 | SO45092 | 313 | 18899 |
| 13 | 13 | 1/3/2015 | 12/16/2001 | SO45089 | 351 | 25977 |
| 14 | 14 | 1/3/2015 | 10/26/2001 | SO45091 | 314 | 18909 |
| 15 | 15 | 1/3/2015 | 9/11/2001 | SO45087 | 350 | 11388 |
| 16 | 16 | 1/3/2015 | 9/11/2001 | SO45094 | 310 | 22785 |
| 17 | 17 | 1/4/2015 | 10/30/2001 | SO45096 | 312 | 12483 |
| 18 | 18 | 1/4/2015 | 10/30/2001 | SO45097 | 313 | 29151 |

HANDS ON INDEX AND CONDITIONAL COLUMN - QUERY EDITOR

1. Create a conditional column that checks if the patient's age is greater than 50. Use 'True' and 'False' to fill in the column.
2. Create a conditional column that checks if the patient's gender is "Male". What is the formula you would use?
3. Create a new Column using conditional column that checks if the patient is "Child " if the age of the child is below 18, "Adult" if the age of the patient is below 65 year of age, "Senior" if the age of the patient is above 65.
4. Create a new Column using conditional column that assigns a value of "High Expense" if the Expenses are greater than 1000, "medium Expense" if the expenses are between 100 -1000 else show "low expense". What is the formula you would use?
5. Create a conditional column that assigns a value of "Neurological Diagnosis" if the DiagnosisCodeGroup contains the word "nervous", otherwise "Non-Neurological Diagnosis". What is the formula you would use?

Answer: if Text.Contains([DiagnosisCodeGroup], "nervous") then "Neurological Diagnosis" else "Non-Neurological Diagnosis"

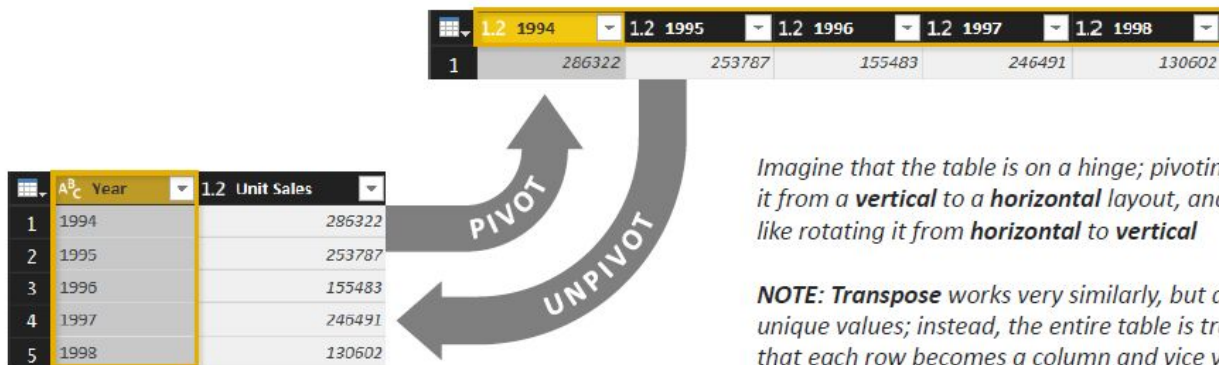
HANDS ON GROUP AND AGGREGATION - QUERY EDITOR

1. Calculate the Total expenses for each Hospital name
2. Which provider has the highest number of visits?
3. What is the total insurance payment received by each provider?
4. How many unique patients are there in the dataset?
Calculate total expenses by each unique patient.
5. Calculate the Expense and Total Payment for each State in a separate table

Note that Group will create a new table with the relevant columns only and remove other columns. To move back to the original table, remove the Group step from applies steps. Alternatively you can create a duplicate copy of the data in the query.

PIVOT AND UNPIVOT - QUERY EDITOR

“Pivoting” is a fancy way to describe the process of turning **distinct row values** into **columns** (“pivoting”) or turning **columns** into **rows** (“unpivoting”)



Imagine that the table is on a hinge; pivoting is like rotating it from a **vertical** to a **horizontal** layout, and unpivoting is like rotating it from **horizontal** to **vertical**

NOTE: *Transpose* works very similarly, but doesn't recognize unique values; instead, the entire table is transformed so that each row becomes a column and vice versa

HANDS ON PIVOT, UNPIVOT AND TRANSPOSE - QUERY EDITOR

1. Connect with 'HC_Pivot_Unpivot_data.csv'.
2. Create a Pivot table from this data to view the CPTUnits for each Hospital and Payer
3. Now use Transpose option to view the table

MERGING QUERIES - QUERY EDITOR

Merge

Select a table and matching columns to create a merged table.

AW_Sales_Data

| OrderDate | ProductKey | CustomerKey | OrderQuantity | StockDate | OrderNumber | TerritoryKey | Order |
|------------|------------|-------------|---------------|------------|-------------|--------------|-------|
| 4/25/2007 | 214 | 14729 | 1 | 4/20/2004 | 5073780 | | 7 |
| 7/26/2008 | 214 | 11242 | 1 | 6/27/2003 | 5051437 | | 38 |
| 12/31/2008 | 214 | 21452 | 1 | 11/27/2003 | 5051128 | | 2 |
| 8/28/2007 | 214 | 22748 | 1 | 4/26/2004 | 5074089 | | 8 |
| 10/6/2009 | 214 | 25635 | 1 | 8/18/2003 | 5055671 | | 4 |

AW_Product_Lookup

| ProductKey | ProductSubcategoryKey | ProductSKU | ProductName | ModelName | ProductDescription |
|------------|-----------------------|------------|-------------------------|---------------------|---------------------|
| 214 | 82 | HL-4509-R | Sport-300 Helmet, Red | Sport-300 | Universal fit, well |
| 215 | 82 | HL-4509 | Sport-300 Helmet, Black | Sport-300 | Universal fit, well |
| 216 | 82 | HL-4509 | Sport-300 Helmet, Black | Sport-300 | Universal fit, well |
| 217 | 82 | HL-4509 | Sport-300 Helmet, Black | Sport-300 | Universal fit, well |
| 218 | 23 | SO-8929-M | Mountain Bike Socks, M | Mountain Bike Socks | Continuation of r |

Join Kind
Left Outer (all from first, matching from second)

✓ The selection has matched 56046 out of the first 56046 rows.

OK Cancel

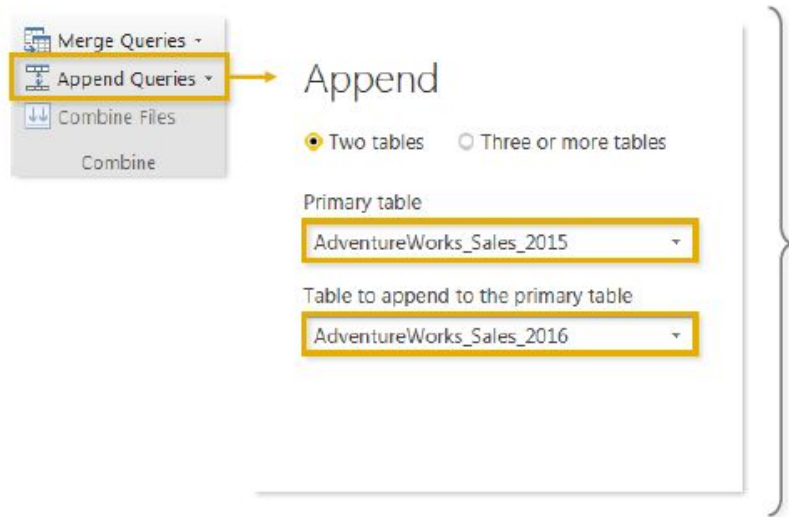
1. Merging queries allows you to **join tables based on a common column** which are usually the Primary Key.
2. Merging queries work exactly like VLOOKUP in excel. Merging adds columns to an existing table

IMPORTANT - Just because you can merge tables, doesn't mean you should. In general, it's better to keep tables separate and define relationships between them.

HANDS ON MERGE - QUERY EDITOR

1. Connect with 'HC_16122019.xlsx' and 'Patient_Details'.
2. Use Merge option to merge both tables.
3. Select 'HC_16122019.xlsx' as primary table

APPENDING QUERIES - QUERY EDITOR

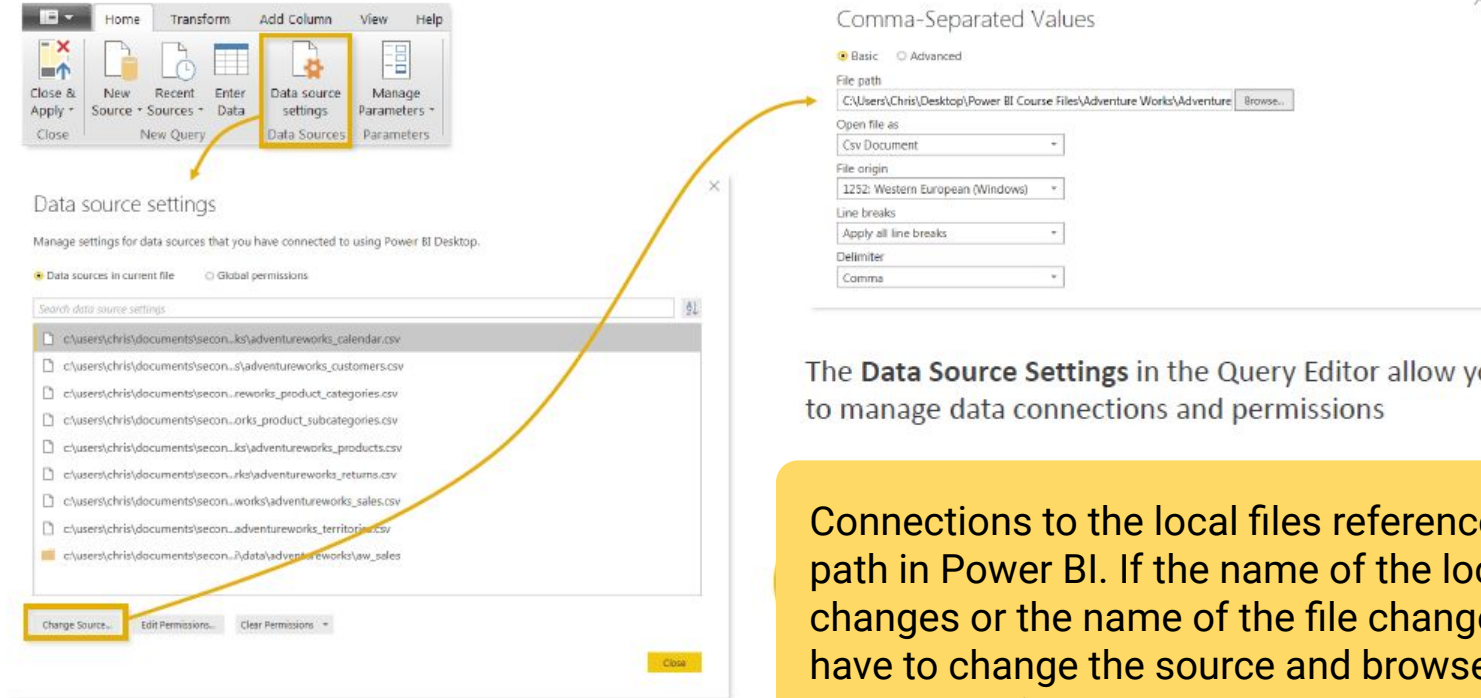


1. Appending queries allow you to combine (or stack) tables which share the same column structure and data types
2. Appending can be done only when data tables -
 - a. have the same columns
 - b. have the same table structures
 - c. have the same data types
3. You can use a '**folder option**' (Get data -> More -> Folder) to append all the files within the folder, if they meet the above conditions. As you add new files, simply refresh the query and they will automatically append.

HANDS ON APPEND - QUERY EDITOR

1. Connect with 'HC_16122019.csv', 'HC_17122019.csv' and 'HC_18122019.csv'.
2. Use append query option to append the 3 tables.
3. Now delete the appended table and the original queries
4. Use folder option to append the 3 tables.

DATA SOURCE SETTINGS



The screenshot illustrates the process of managing data source settings in the Power BI Query Editor. The 'Data source settings' button in the ribbon is highlighted, leading to the 'Data source settings' dialog box. This dialog allows users to manage settings for data sources connected to the Power BI Desktop. The 'Data sources in current file' tab is selected, displaying a list of CSV files. The 'Change Source...' button is highlighted, leading to the 'Comma-Separated Values' dialog box, which shows the 'Basic' tab with fields for File path, Open file as, File origin, Line breaks, and Delimiter.

The **Data Source Settings** in the Query Editor allow you to manage data connections and permissions

Connections to the local files reference the exact path in Power BI. If the name of the location changes or the name of the file changes, you will have to change the source and browse to the current version

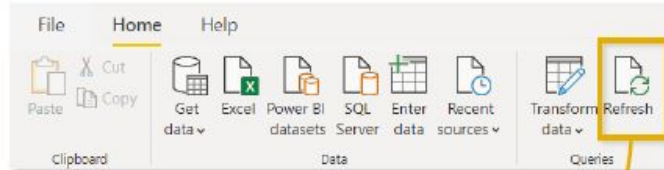
DATA SOURCE SETTINGS - QUERY EDITOR

1. In the Power Query, click on the Home -> Data Source Settings
2. This Data Source Settings window shows the path for each of the files
3. Lets us tentatively change the name of 'HC_16122019.csv' to "HC_16122020'.csv Now go to Home in 'Power Query' and click on 'Refresh preview' icon. An error message come up indicating a Data Source Error



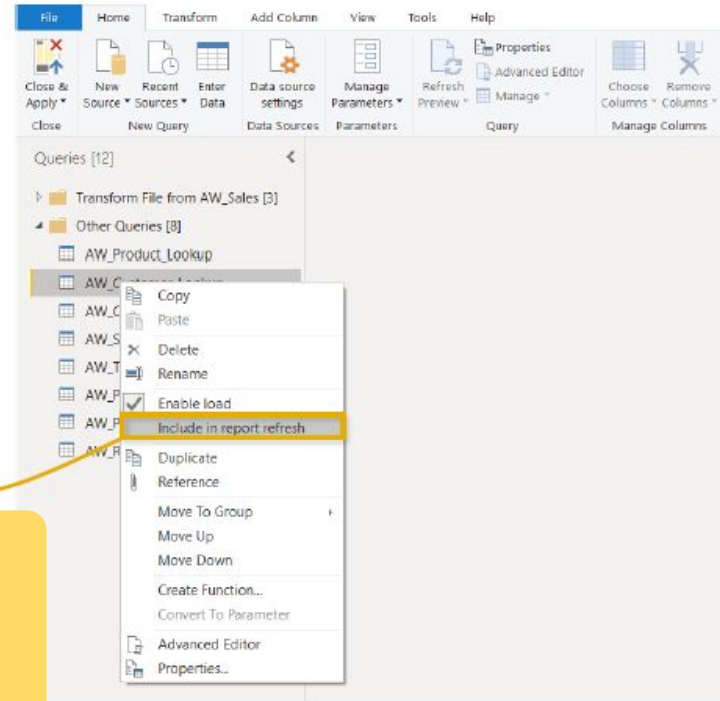
1. Again Go to the Power Query, click on the Home -> Data Source Settings
2. Update the path using 'Change Source' in the Data Source settings
3. Close and Refresh Preview in the Query Editor

REFRESH QUERIES



By default, **ALL** queries in the model will refresh when you use the “*Refresh*” command from the **Home** tab

From the Query Editor, uncheck “***Include in report refresh***” to exclude individual queries from the refresh



Avoid refreshing queries that do not change often to do not change at all by opening the Power Query mode and clicking on the Queries (in the left) and unchecking the 'Include in report refresh' option.