Power BI training

Power BI

Turn your data into Impact!

INTRODUCTION TO CASE

Power BI

THE SITUATION

You are hired by the CEO of health care Industry. to deep dive into their reports can create a comprehensive Business Intelligence solution. As an data analyst you are given access to the reports

THE BRIEF

Ceo wants you to track critical KPI's (Sales, Revenues, Profits and Returns), compare regional performance, analyse product level trends and forecasts

THE OBJECTIVE

Use Power BI Desktop to -

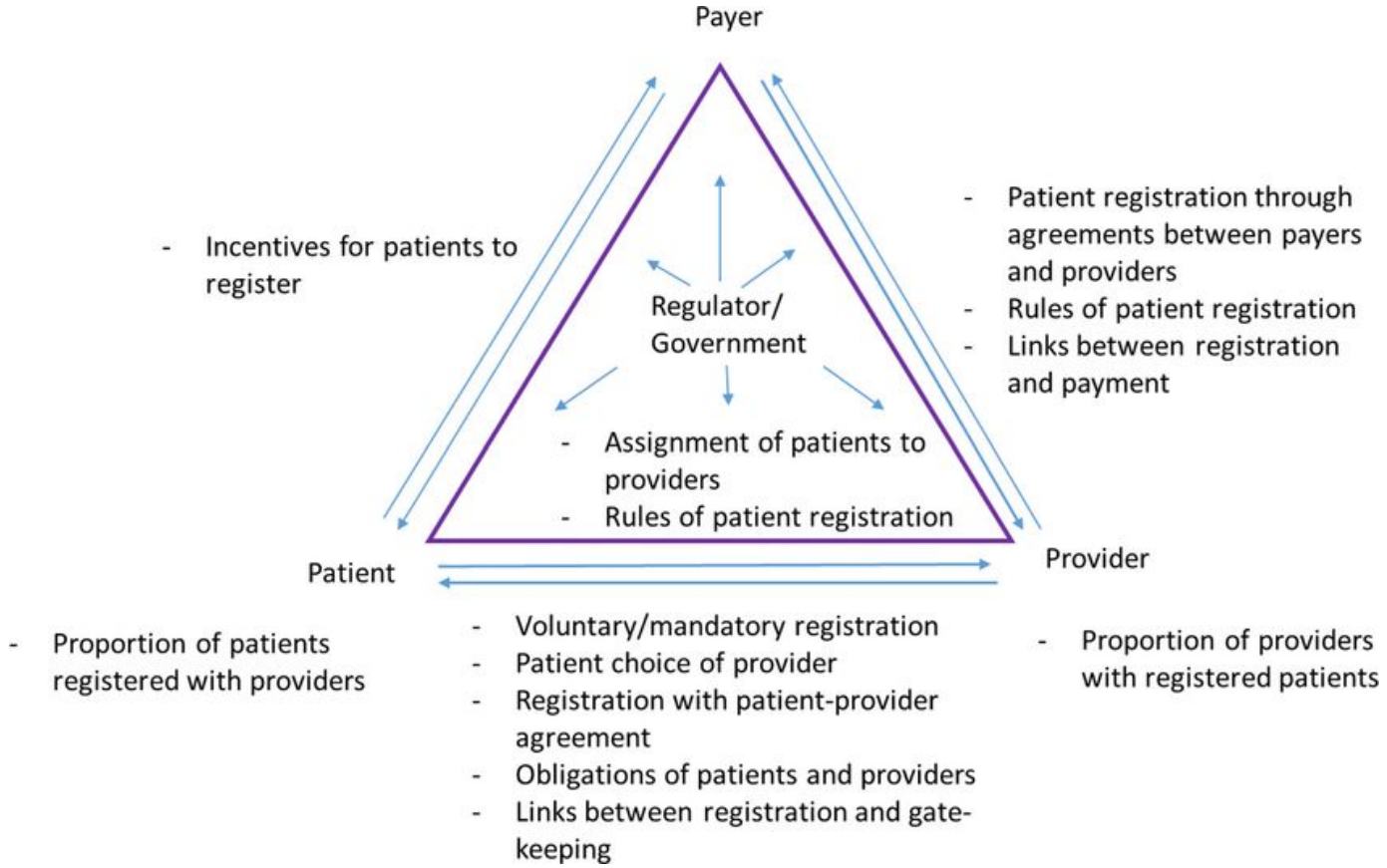
- Connect and transform the raw data
- Build a relational data model
- Create new calculated columns and measures using DAX
- Design an interactive report to analyse and visualize the data

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



About the dataset - data dictionary

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.

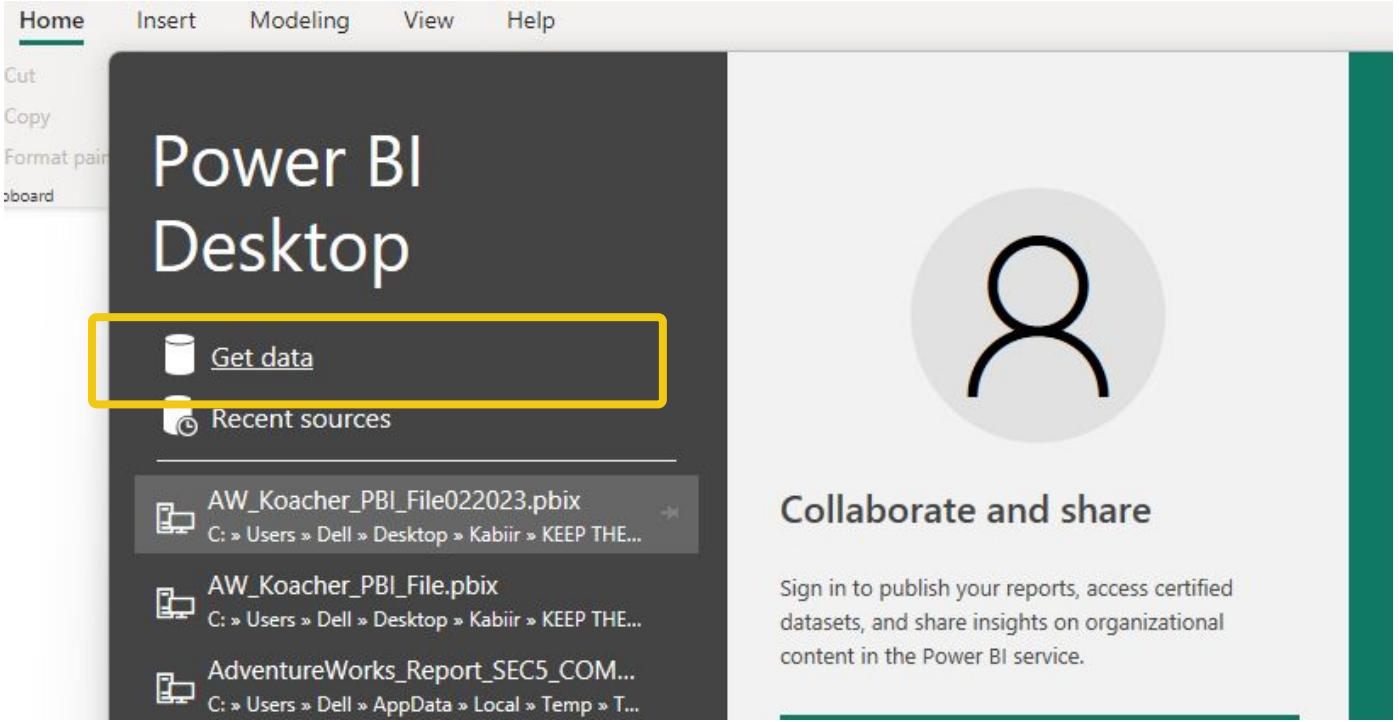
About the dataset - data dictionary

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 & SHAPING DATA

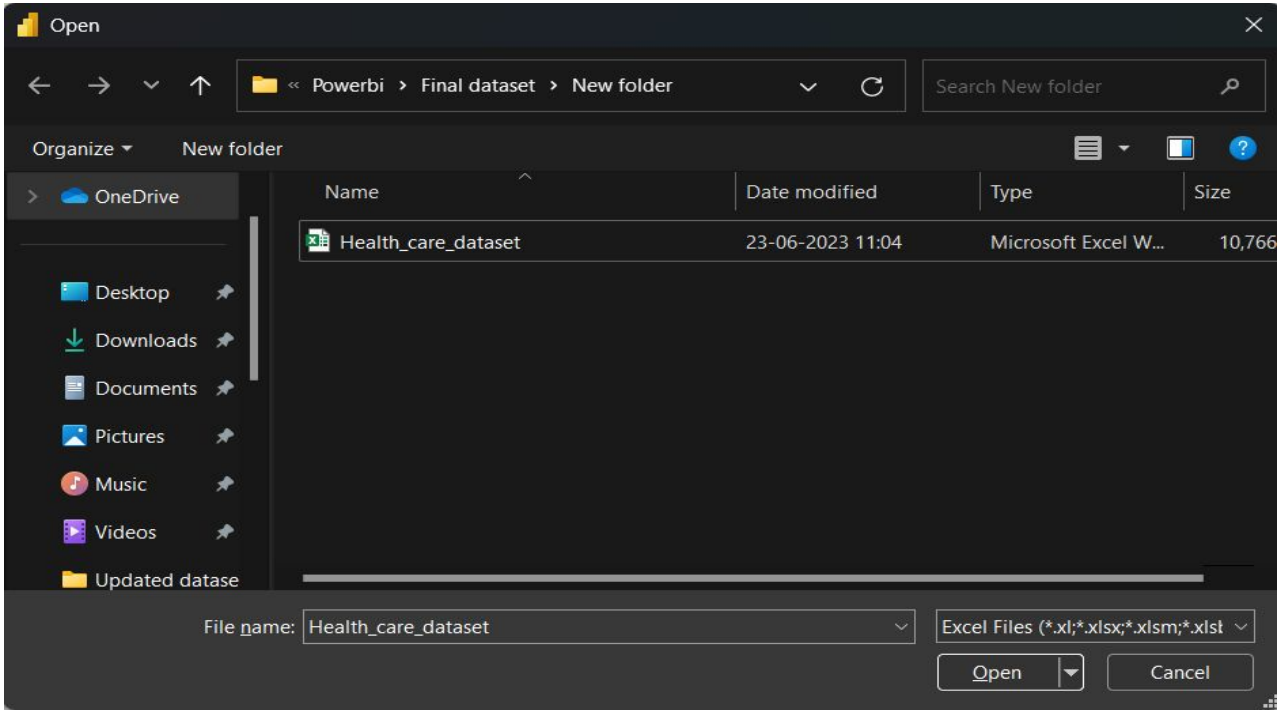
HANDS ON - CONNECTING DATA

1. Connect with ‘health_care_dataset.csv’ file and go to

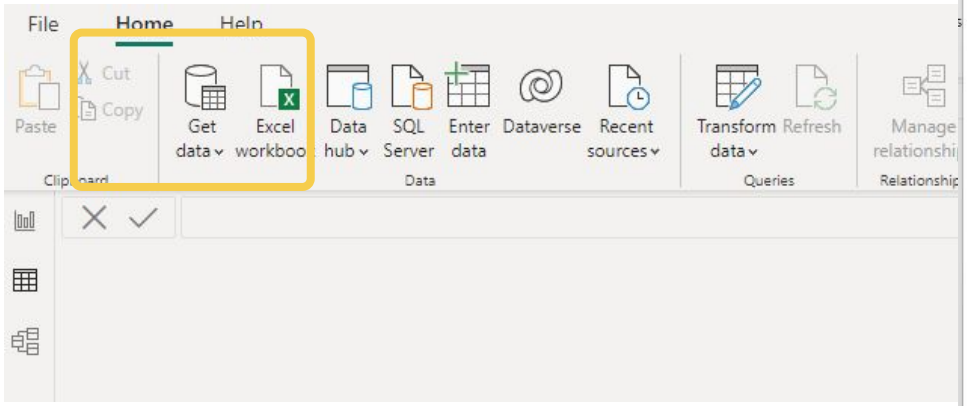


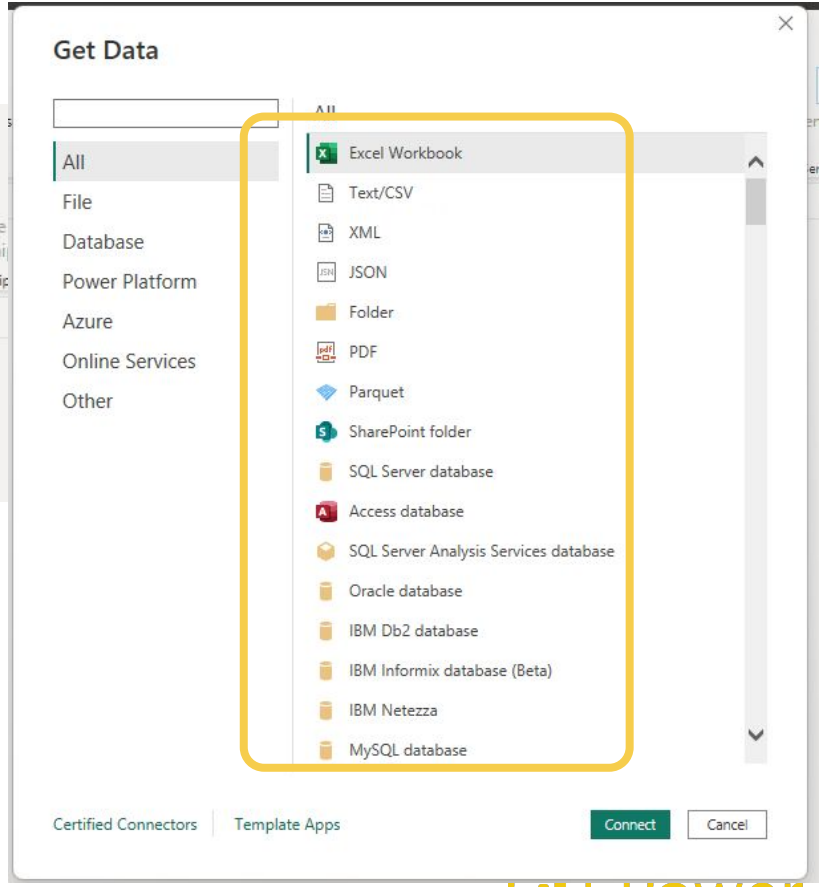
HANDS ON - CONNECTING DATA

1. Connect with ‘health_care_dataset.csv’ file and go to



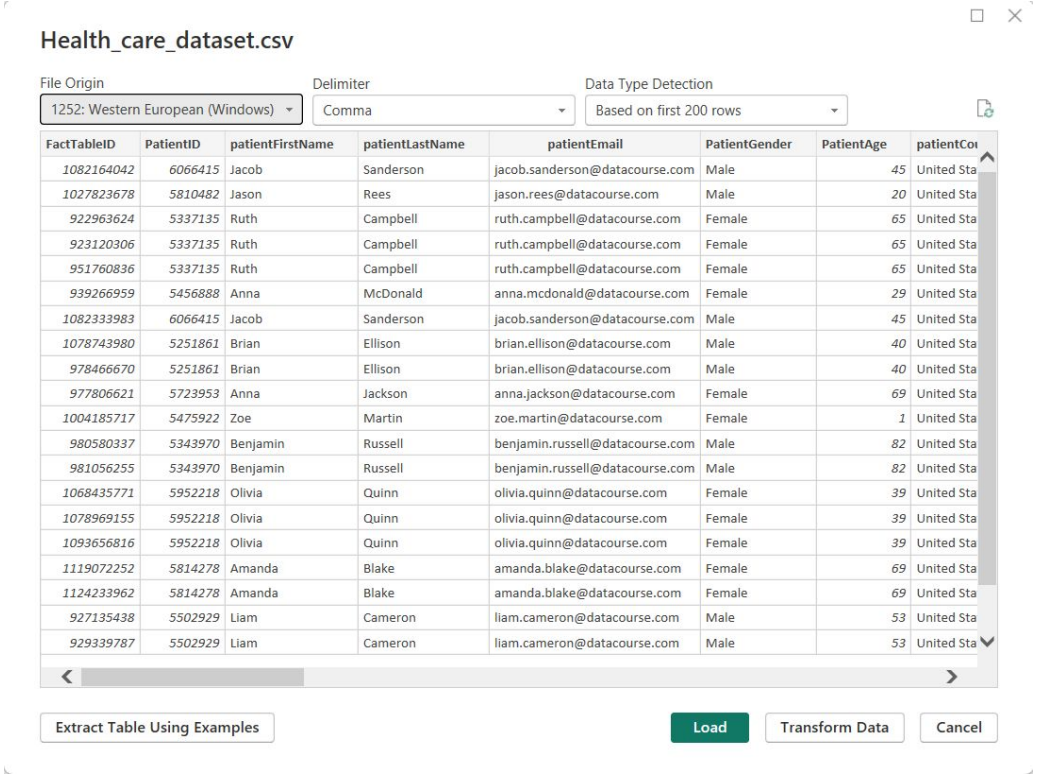
Connect to data from Home tab





HANDS ON - QUERY EDITOR

1. Connect with ‘health_care_dataset.csv’ file and go to

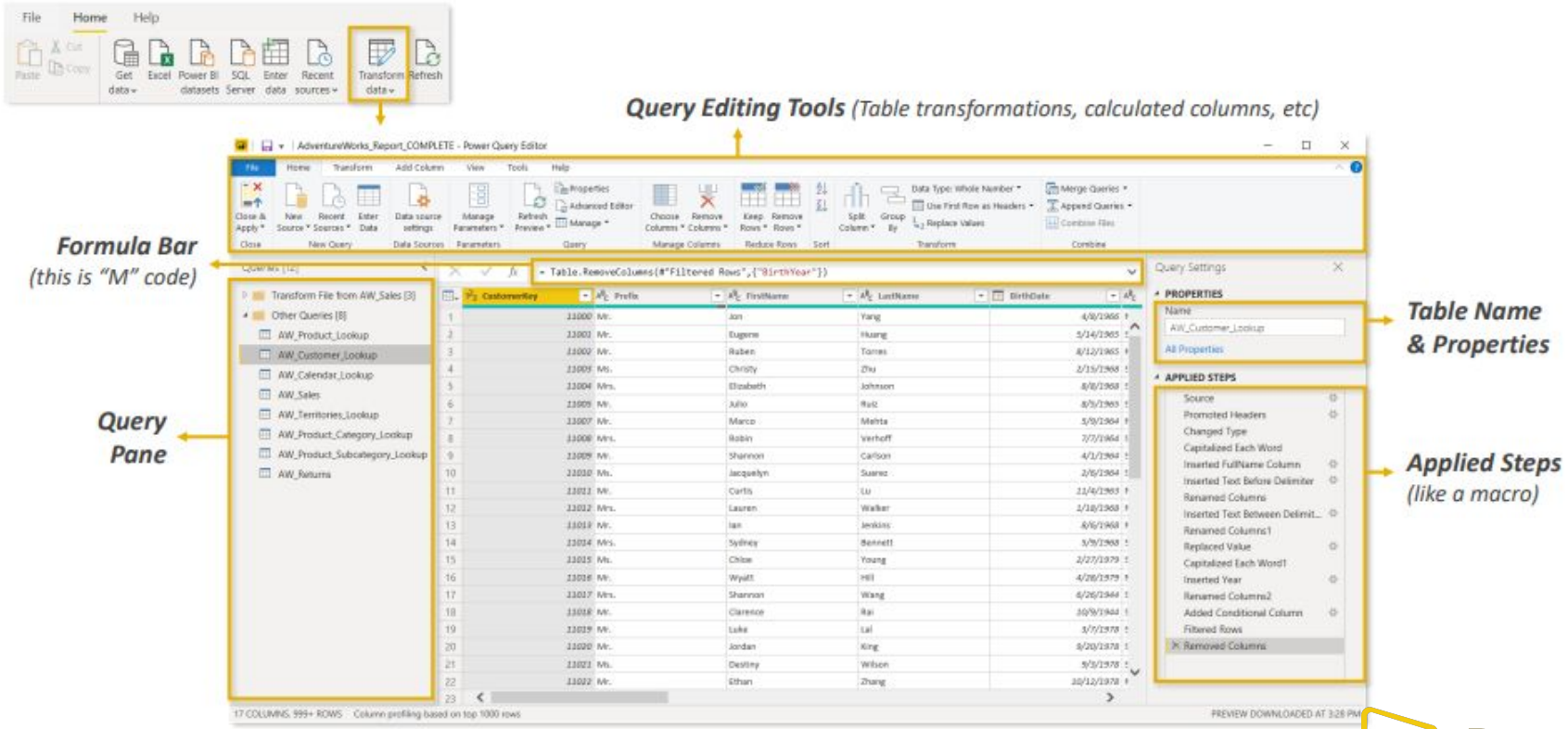


FactTableID	PatientID	patientFirstName	patientLastName	patientEmail	PatientGender	PatientAge	patientCou
1082164042	6066415	Jacob	Sanderson	jacob.sanderson@datacourse.com	Male	45	United Sta
1027823678	5810482	Jason	Rees	jason.rees@datacourse.com	Male	20	United Sta
922963624	5337135	Ruth	Campbell	ruth.campbell@datacourse.com	Female	65	United Sta
923120306	5337135	Ruth	Campbell	ruth.campbell@datacourse.com	Female	65	United Sta
951760836	5337135	Ruth	Campbell	ruth.campbell@datacourse.com	Female	65	United Sta
939266959	5456888	Anna	McDonald	anna.mcdonald@datacourse.com	Female	29	United Sta
1082333983	6066415	Jacob	Sanderson	jacob.sanderson@datacourse.com	Male	45	United Sta
1078743980	5251861	Brian	Ellison	brian.ellison@datacourse.com	Male	40	United Sta
978466670	5251861	Brian	Ellison	brian.ellison@datacourse.com	Male	40	United Sta
977806621	5723953	Anna	Jackson	anna.jackson@datacourse.com	Female	69	United Sta
1004185717	5475922	Zoe	Martin	zoe.martin@datacourse.com	Female	1	United Sta
980580337	5343970	Benjamin	Russell	benjamin.russell@datacourse.com	Male	82	United Sta
981056255	5343970	Benjamin	Russell	benjamin.russell@datacourse.com	Male	82	United Sta
1068435771	5952218	Olivia	Quinn	olivia.quinn@datacourse.com	Female	39	United Sta
1078969155	5952218	Olivia	Quinn	olivia.quinn@datacourse.com	Female	39	United Sta
1093656816	5952218	Olivia	Quinn	olivia.quinn@datacourse.com	Female	39	United Sta
1119072252	5814278	Amanda	Blake	amanda.blake@datacourse.com	Female	69	United Sta
1124233962	5814278	Amanda	Blake	amanda.blake@datacourse.com	Female	69	United Sta
927135438	5502929	Liam	Cameron	liam.cameron@datacourse.com	Male	53	United Sta
929339787	5502929	Liam	Cameron	liam.cameron@datacourse.com	Male	53	United Sta

Clicking on the ‘Load’ will load the data into the main interface where you view, create models and build visualizations.
Clicking on the ‘Transform Data’ will open a Query Editor window.

Click on the ‘Transform Data’ to open the Query Editor window.

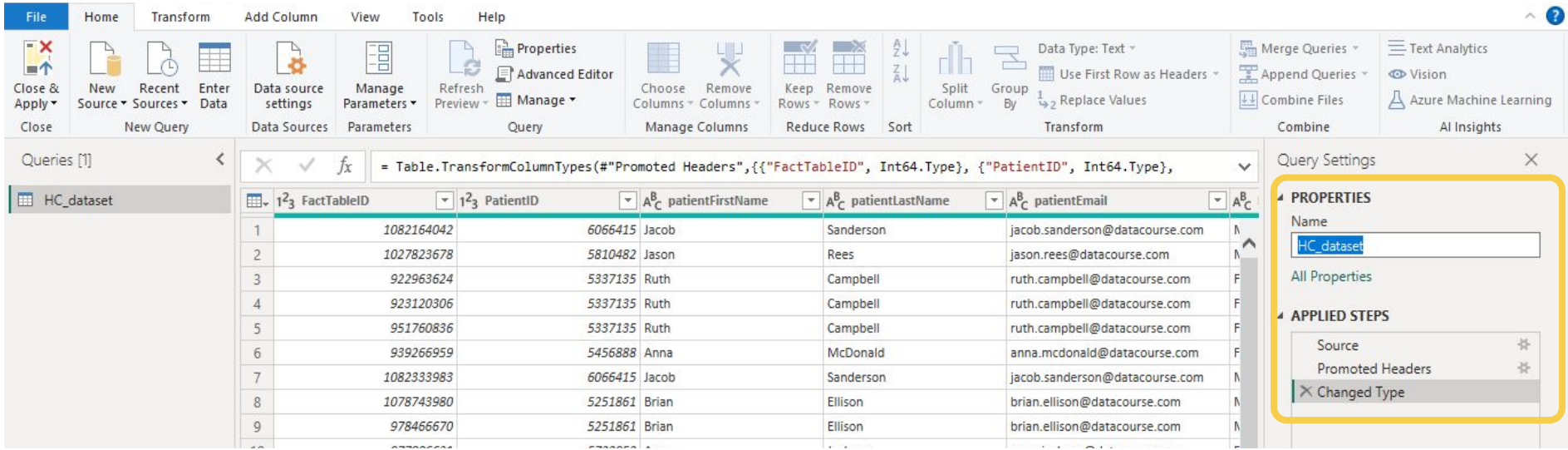
THE QUERY EDITOR



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

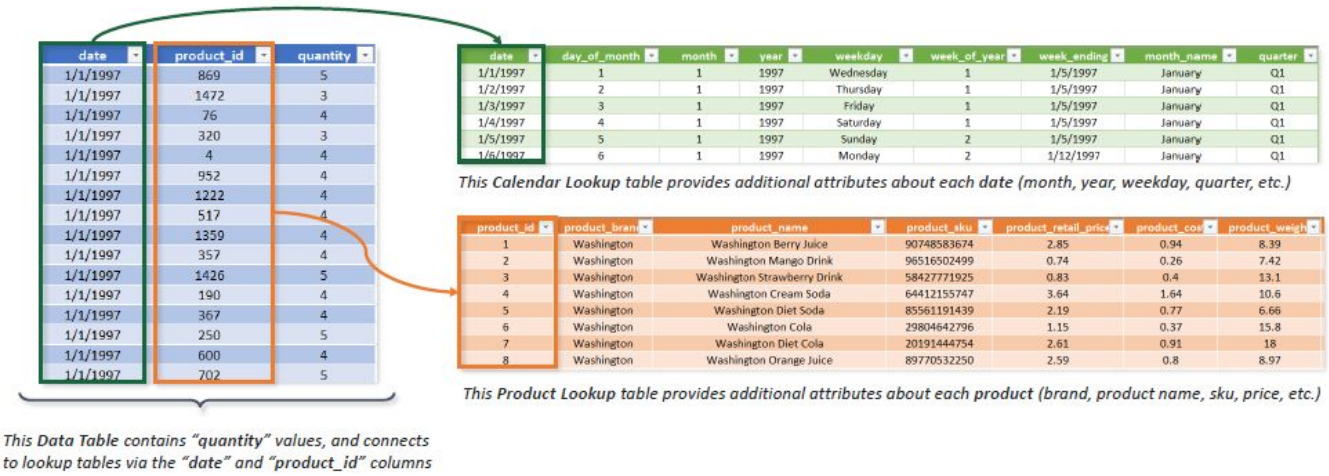


DATA TABLES AND LOOK-UP TABLES

Data Model contains 2 types of tables -

Data Tables (are fact tables) - contains numbers and values. They are granular tables with 'ID' or 'Key' fields which is used to create relationship between them and the 'Look up Tables' e.g. Sales table,, Returns table

Look Up Tables contains descriptions and details which are usually text based attributes about dimensions in the data e.g - Customer, Products, 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.)

Power BI

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_[illegible]	product_name	product_sku	product_retail_price	product_cost	product_weigh
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

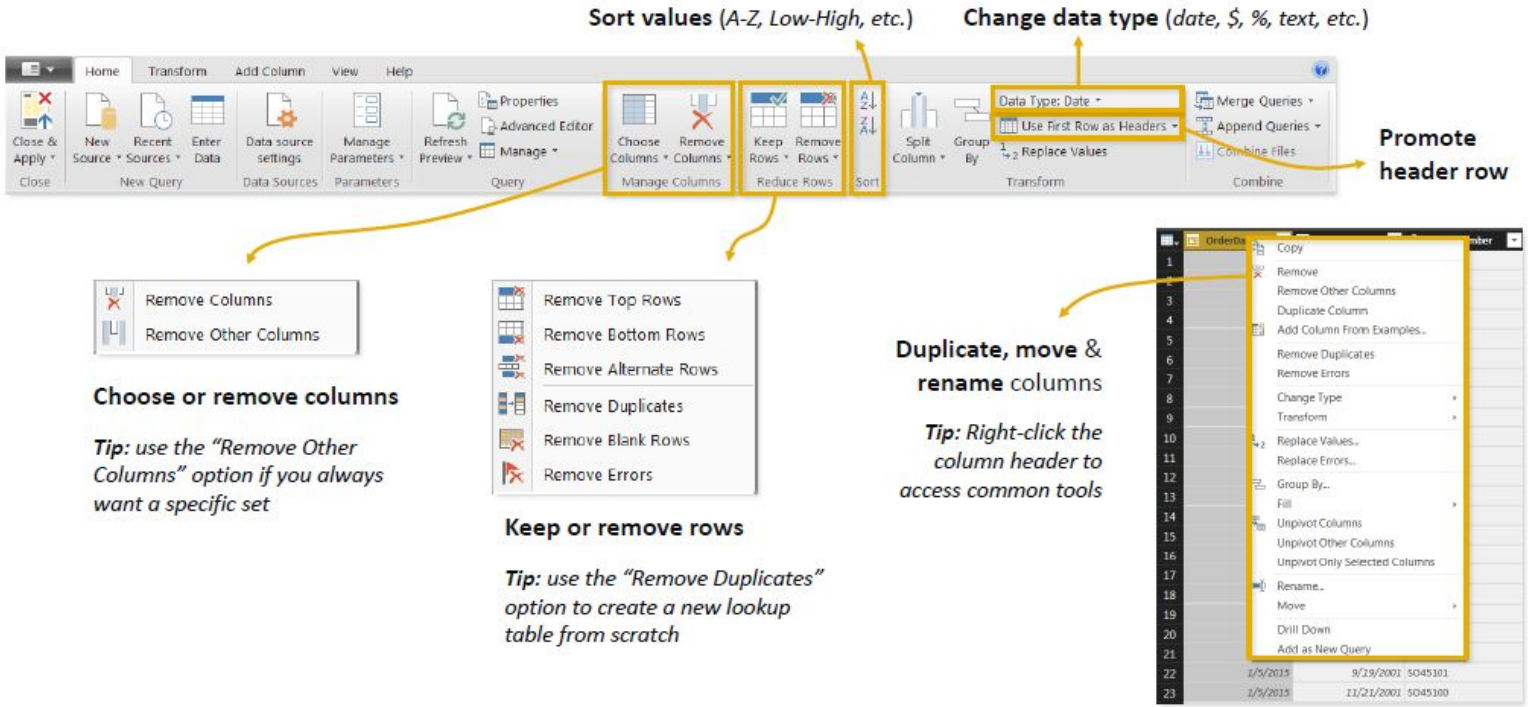
Foreign keys exist usually in Data tables*. They contain multiple instances and are mapped with the Primary keys from the 'Look up Tables'

Primary keys exist usually in 'Lookup tables'*. They are Unique and are used with Foreign Keys to map the attributes/ details of the dimension

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

Power BI

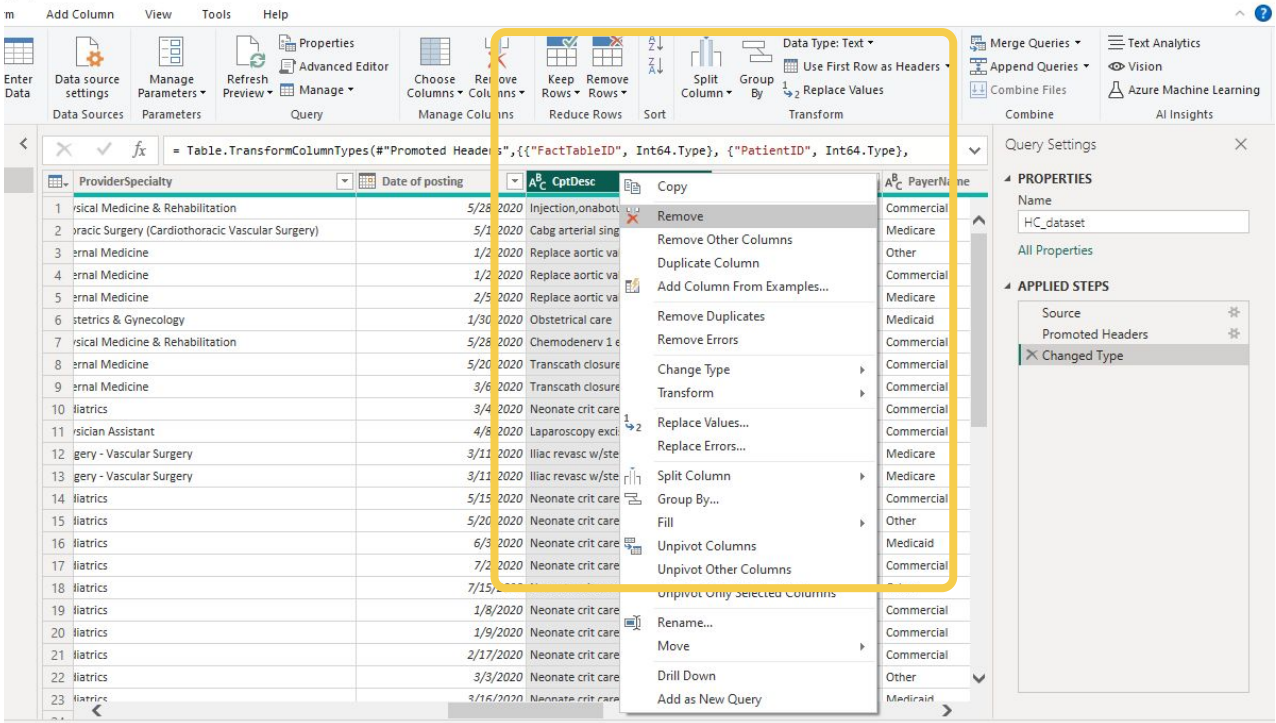
QUERY EDITOR - BASIC TRANSFORMATIONS



Power BI

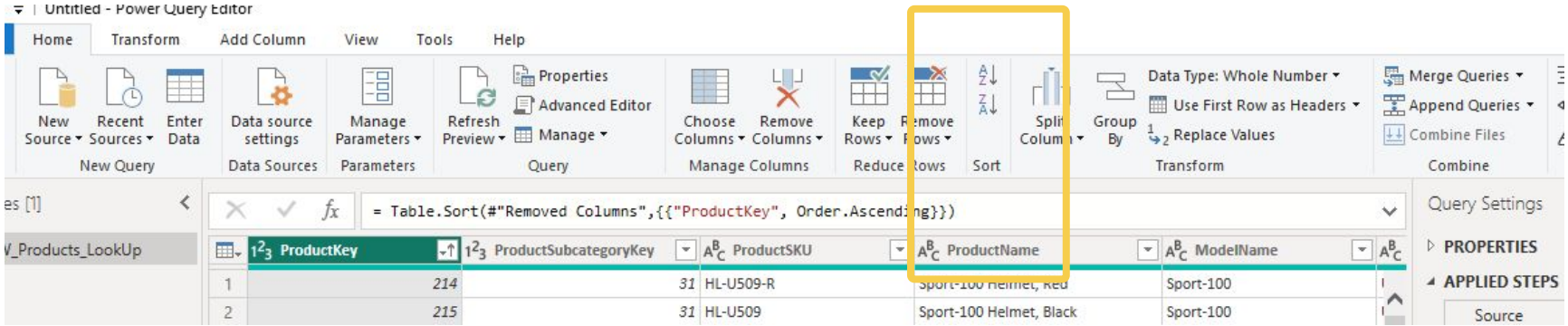
HANDS ON - QUERY EDITOR

Click on **'CptDesc'** and remove this column



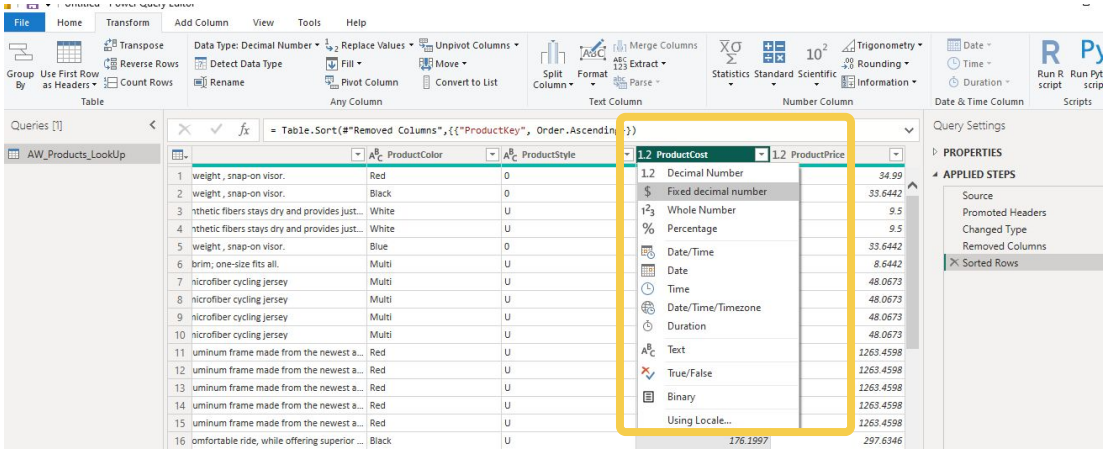
HANDS ON - QUERY EDITOR

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



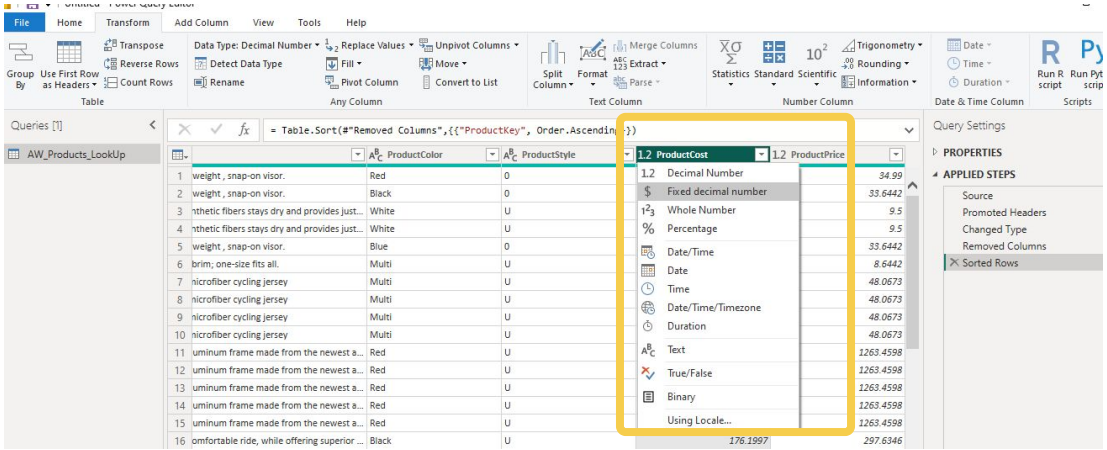
HANDS ON - QUERY EDITOR

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.
3. Hit ‘Close and Apply’



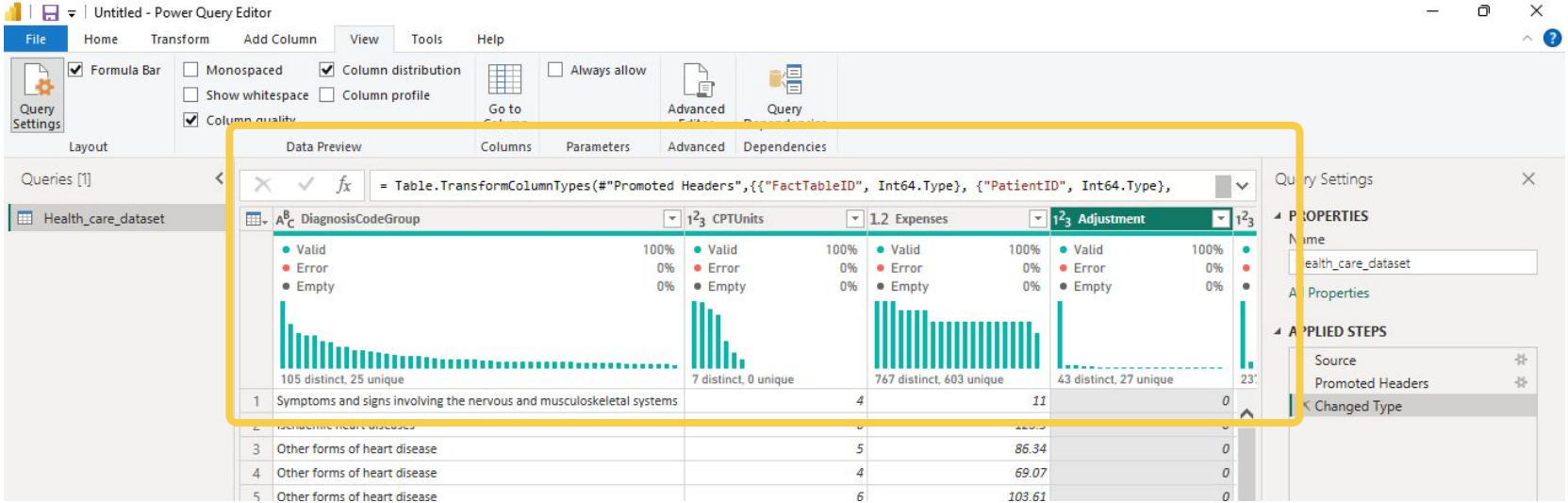
HANDS ON - QUERY EDITOR

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.
3. Go to View and check for the Column profiles.



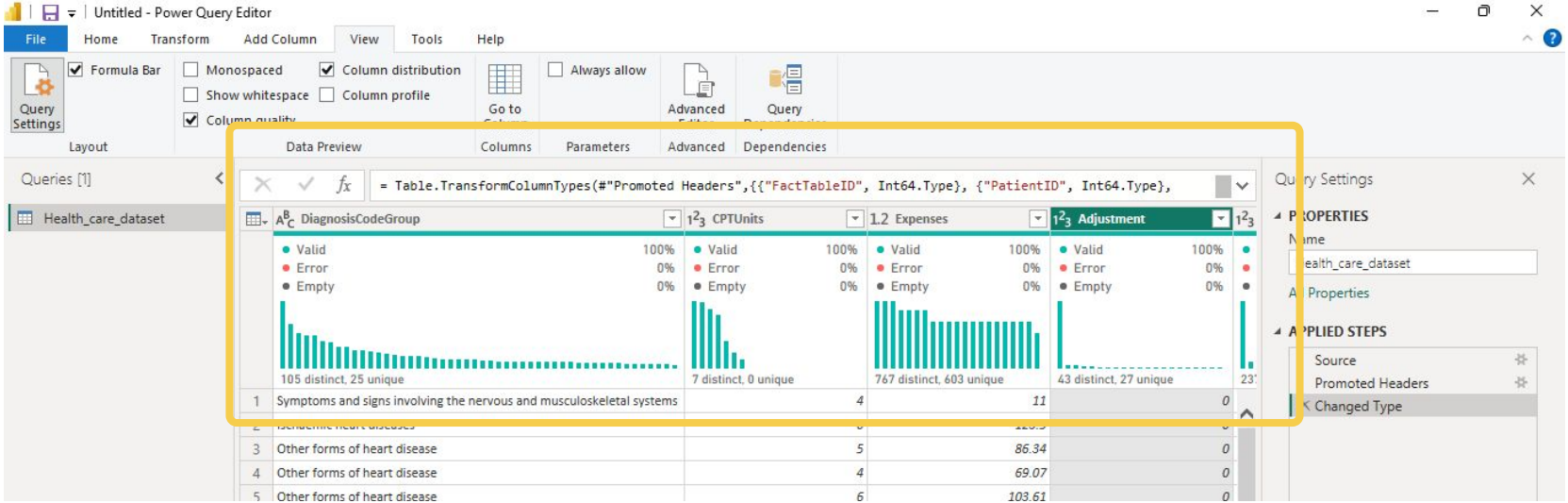
HANDS ON - COLUMN PROFILES

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

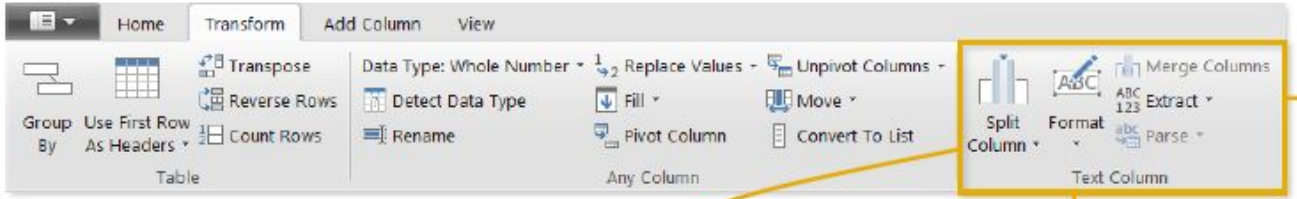


HANDS ON - TREATMENT OF ERRORS

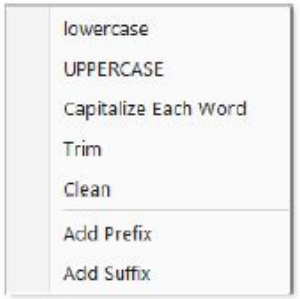
1. Make sure the column profiling is done using entire dataset
2. Now In the ‘Adjustment’ column, view the rows having errors. Analyse if these are required to be kept or deleted.
3. Keep the errors and replace the Error value by ‘0’.



QUERY EDITOR - TEXT TRANSFORMATIONS

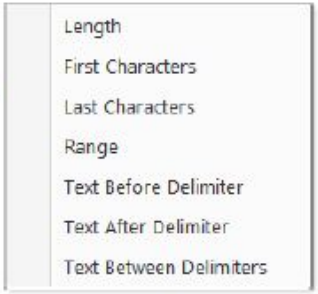


Split a text column based on either a specific delimiter or a number of characters



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*



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*

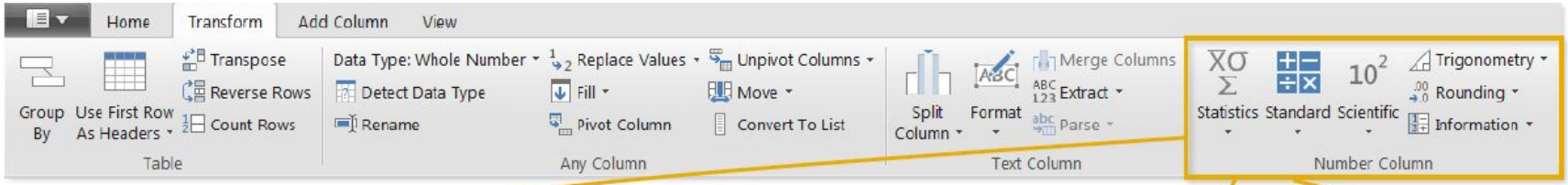
'Transform' and 'Add Column' have many similar tools.
'Transform' tab modifies the existing columns using the tools and 'Add Column' adds a new column by applying the tools.

HANDS ON TEXT TRANSFORMATIONS - QUERY EDITOR

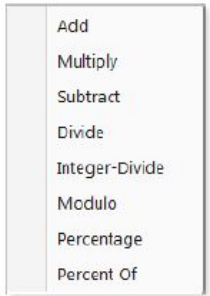
1. Connect a new data table - health_care_dataset.csv.
2. Load into the Query Editor
3. Perform the following transformations -
    a. Extract the patient's email domain from the "patientEmail" column
    b. Capitalize the first letter of each word in the "patientFirstName" column
    c. Concatenate the patient's first name and last name into a single column
    d. Calculate the length of each patient's email address in the "patientEmail" column
    e. Replace a specific word (e.g., "Cramp") in the "DiagnosisCodeDescription" column with another word (e.g., "Pain")?
    f. Extract the first 3 letters from patientState and create a new variable - ‘State Code’
    g. Similarly create a Country Code column and put value USA in it

Power BI

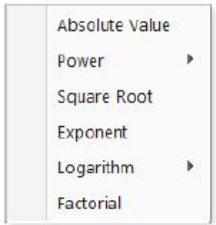
QUERY EDITOR - NUMERIC TRANSFORMATIONS



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



Standard



Scientific

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

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 returns a SINGLE value and are generally used to explore data while Standard, Scientific and Trigonometry tools apply to each row.

Power BI

HANDS ON NUMERIC TRANSFORMATIONS - QUERY EDITOR

1. Calculate the average age of the patients.
2. Find the maximum age among the patients.
3. Subtract 10 from the PatientAge field.
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 ('Insurance_Payment' + 'Patient_Payment')/Expenses. Apply the unit % to the column.

Power BI

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*

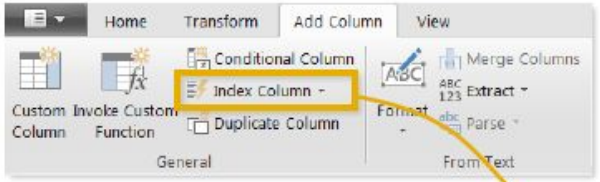
Power BI

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?

Power BI

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

Power BI

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"

Power BI

HANDS ON GROUP AND AGGREGATION - QUERY EDITOR

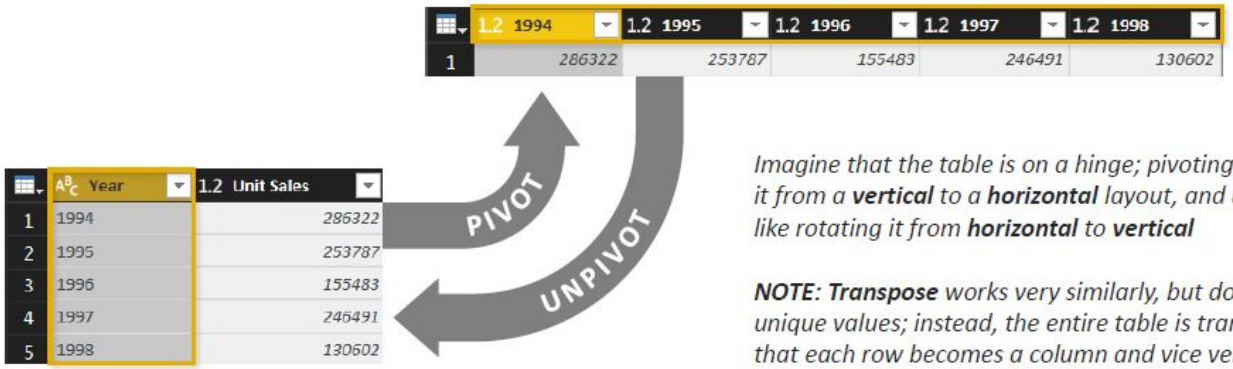
1. Calculate the Total expenses for each Hospital name
2. Which provider has the highest number of visits?
3. What is the total insurance payment received by each provider?
4. How many unique patients are there in the dataset? Calculate total expenses by each unique patient.
5. Calculate the Expense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

Power BI

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*

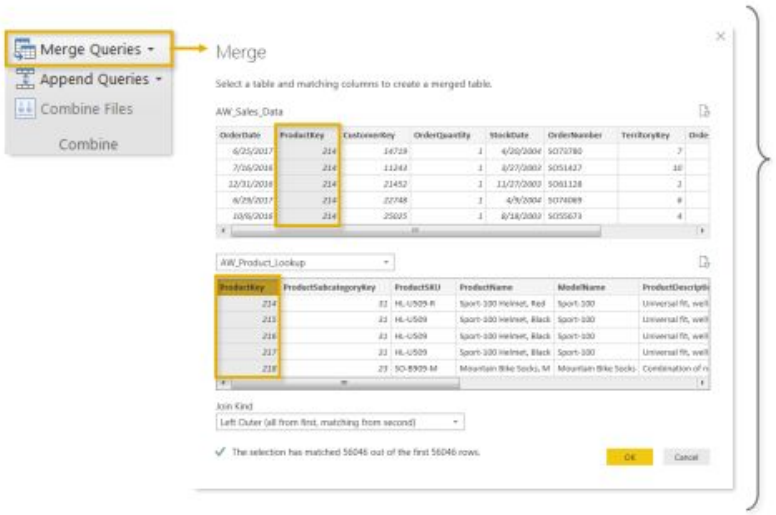
Power BI

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

Power BI

MERGING QUERIES - QUERY EDITOR



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.

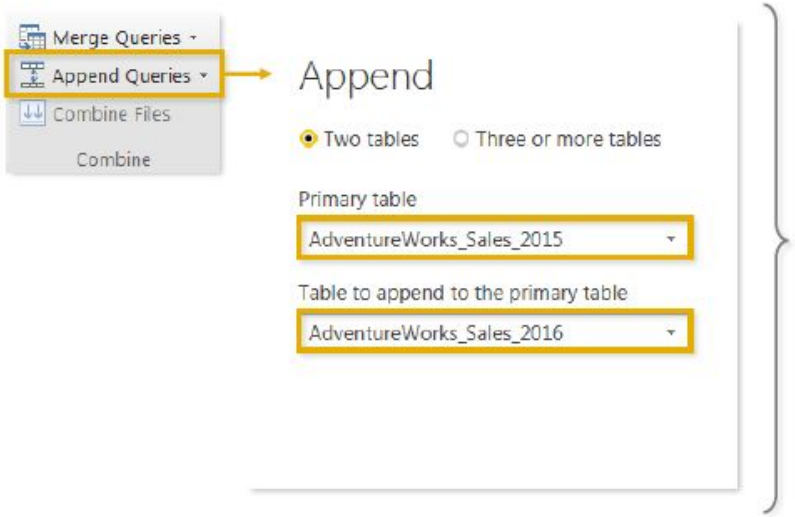
Power BI

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

Power BI

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.

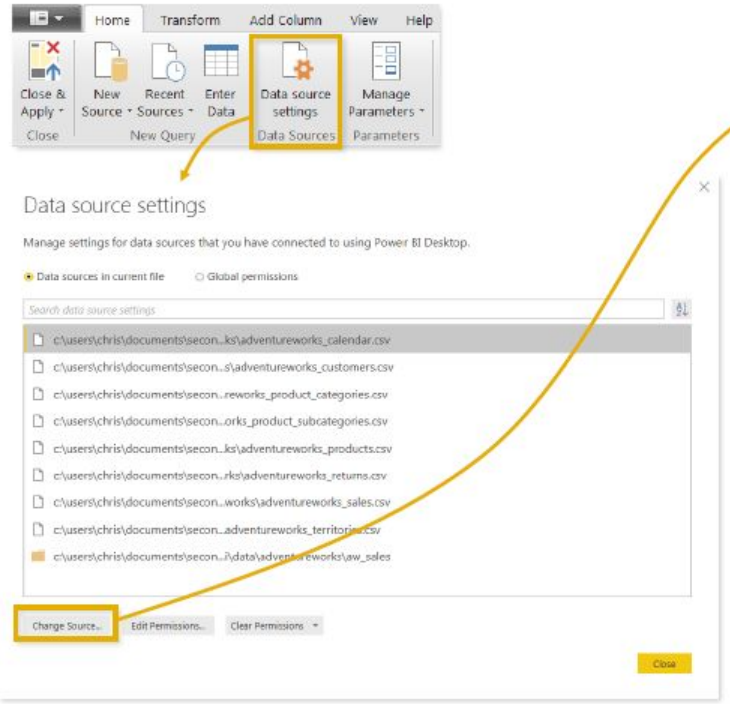
Power BI

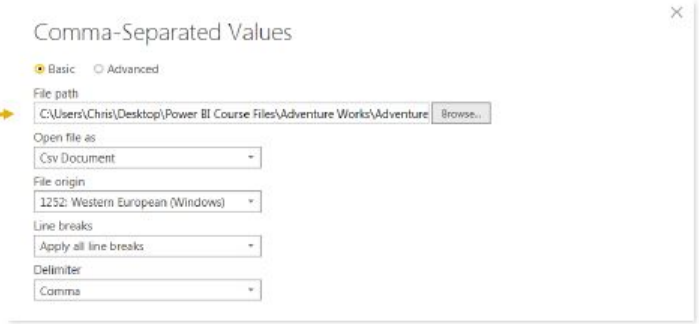
HANDS ON APPEND - QUERY EDITOR

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.

Power BI

DATA SOURCE SETTINGS





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

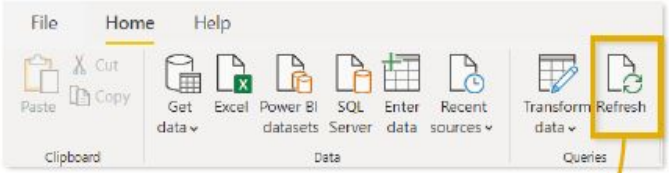
Connections to the local files reference the exact path in Power BI. If the name of the location changes or the name of the file changes, you will have to change the source and browse to the current version

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

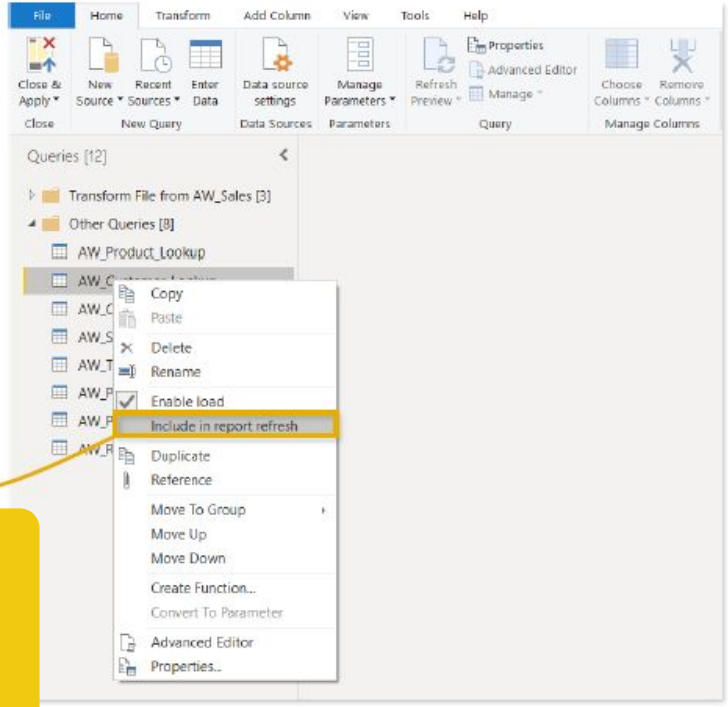
1. Again Go to the Power Query, click on the Home -> Data Source Settings
2. Update the path using 'Change Source' in the Data Source settings
3. Close and Refresh Preview in the Query Editor

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.