MMPA

Day 2

Data Analytics

MGT 1272 – MANAGEMENT INFORMATION SYSTEMS

Behram Faroogh

What will we learn today?

1. Data Analytics
2. Data Visualization
3. Data Wrangling
4. Power Query

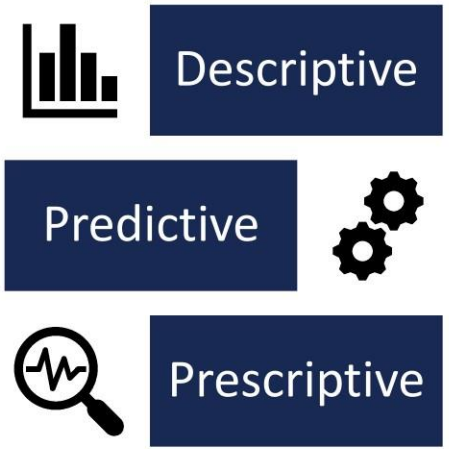
1. Data Analytics

What is Analytics?

Analytics is an encompassing, multidimensional field that uses mathematics, statistics, predictive modelling and machine learning techniques to find meaningful information from data. It is the discovery, interpretation and communication of meaningful patterns and trends in data and applying them towards effective and data-driven decision making.

Types of Analytics:

- Descriptive: Analytics that helps you understand how things are going.
  Think: what has happened? *Examples are:* cash flow and revenue analysis.
- Predictive: Analytics that help you forecast future performance and results.
  Think: what could happen? *Example is:* sentiment analysis.
- Prescriptive: Analytics that suggest a prescribed next step or action.
  Think: what should we do? *Example is:* Google's self-driving car.



Data Analytics is the science of examining raw data in order to draw conclusions about the information it provides.

Data Analytics

Impact on Business

- **enables** new products and services which can lead to the creation and emergence of new markets. An example of new products includes wearables such as Fitbits or smart jewelry.
- **disrupts** existing markets and unseats traditionally secured businesses through innovation. Uber is a good example where data analytics has been a major factor in the growth of the company and disruption of the taxi industry.
- **drives** increased efficiency. For example, with the help of data analytics, retailers are able to optimize supply chain through custom-tailored offerings based on trends and forecasts that are backed up with data.

CSF for Data Analytics

Purpose

Understanding the goal and the questions that need to be answered is the first step for a successful analysis. Work done with the data should link directly with the issue on hand and should aim to add value to the users.

Know-how

Having the correct level of skills, techniques and judgement is important to be able to collect valuable data efficiently, analyze it to quickly generate outputs and accept, reject or refine hypotheses.

Compelling insight

Skilled analysis develops ideas and reveals insights that may not be obvious. However, it is key to present insights as a visually compelling story that makes complex issues easy to understand.

2. Data Visualization

Data Visualization

Data visualization is the graphical representation of information and data.

It is using visual elements such as graphs, charts, dashboards and maps as well as tools to provide accessible ways to see and understand trends, outliers and patterns. Data visualization is particularly important in the world of Big Data due to the need to analyze massive amounts of information in order to make key decisions.

Quality traits of good data visualization:

- Truthful: avoids deception and dishonesty with audience and maintains clarity when presenting information.
- Functional: built in a way that allows people interpret it correctly and do meaningful operations; it shows connections between data and tells an easy to follow story to allow the audience to quickly understand and use the information presented.
- Beautiful: has an appealing structure, formatting and visual form.
- Insightful: reveals evidence that audience would have a hard time seeing otherwise and clears the path to make valuable discoveries that would be inaccessible if information were presented in a different way.
- Enlightening: changing people's mind for the better and revealing information that increases their well-being.

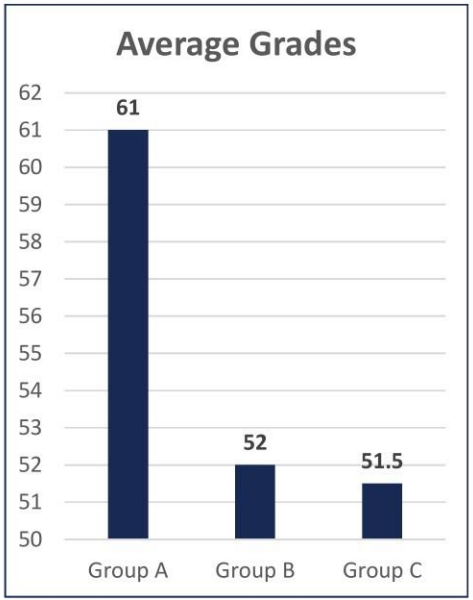
Examples of Misleading Visualization

a. Omitting the baseline

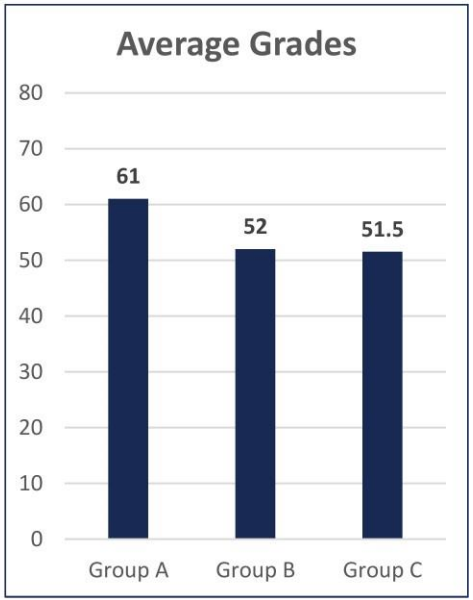
Typically, graphs start with a baseline of 0. However, designers can skew how the data is perceived by changing the baseline to a different number. This is known as a “truncated graph”.

The graph on the left is misleading because the Y-axis starts with a value of 50, making any small difference between the groups seem massive. Group A appears to have a much larger average grade than Groups B and C.

On the other hand, the graph on the right starts with a value of 0 on the Y-axis, offering a more accurate depiction of the data and a less dramatic difference between the groups.



Misleading ✗



Accurate ✓

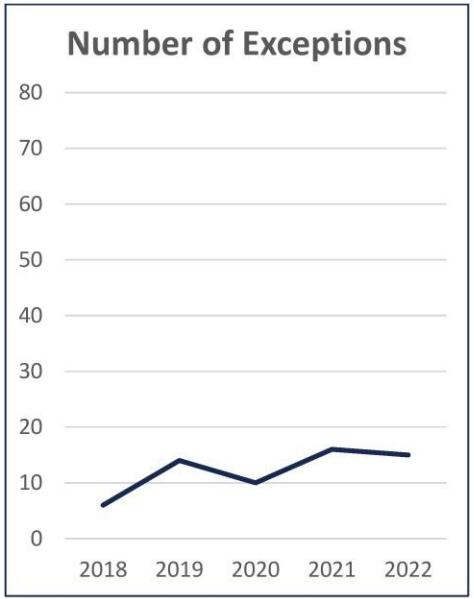
Examples of Misleading Visualization

b. Manipulating the Y-axis

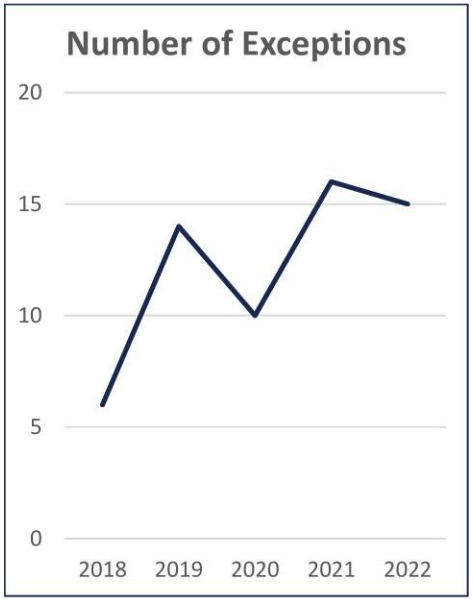
Expanding or compressing the scale on graphs may result in making data seem more or less significant than it actually is.

The graph on the left is misleading because the scale is disproportionate to the data, making the change in the number of exceptions over time seem small.

On the other hand, the graph on the right has a more proportionate scale to the data, and therefore shows a greater change in the number of exceptions over time.



Misleading ✗



Accurate ✓

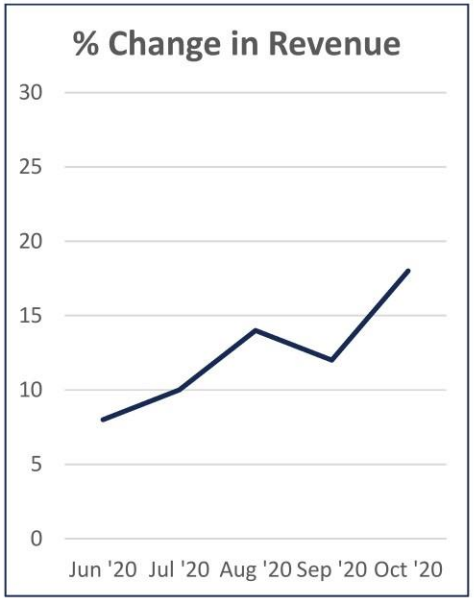
Examples of Misleading Visualization

c. Cherry-picking the data

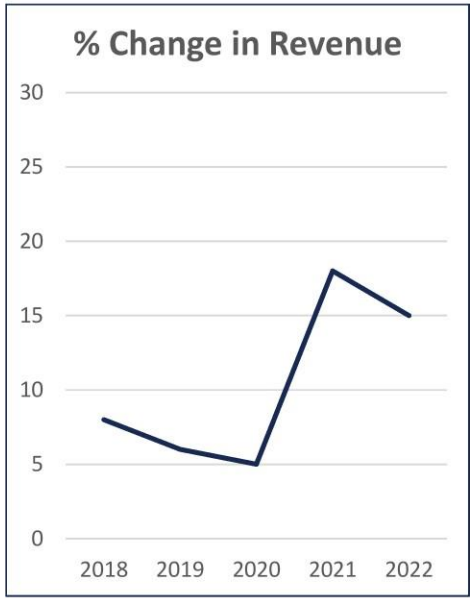
Only certain data points may be included on graphs by designers to reinforce a certain narrative or hidden agenda, leading to a false impression of the data to be perceived.

The graph on the left is misleading because it includes only a few months from the year, depicting an upward percentage increase in revenue.

On the other hand, the graph on the right is more accurate since it includes a wider range of data, revealing the true overall trend of revenue over the years and showcasing the bigger picture.



Misleading ✗



Accurate ✓

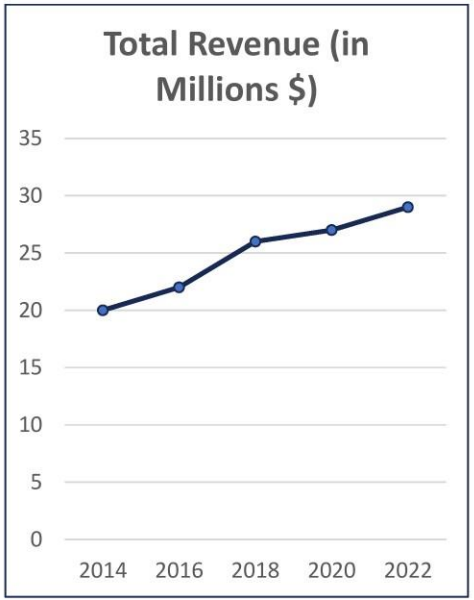
Examples of Misleading Visualization

d. Omitting the data

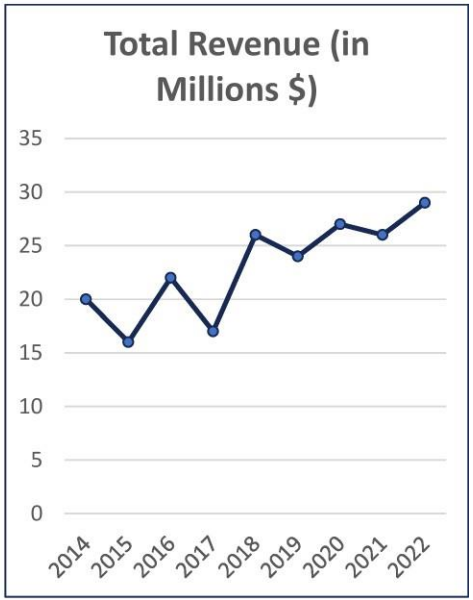
Omitting certain data points allow designers to create trends that don't really exist. This is due to the fact that by omitting data, designers do not present the full context. Simply put, why lie when you can just omit?

The graph on the left shows the total revenue on a bi-annual basis, leaving out some data points, and making the graph look much smoother and stable. Users may perceive a steady increase of revenue over the years with little context.

The graph on the right is more accurate since it plots total revenue for every year, depicting a more accurate representation of a more volatile increase in revenue over the years.



Misleading ✗



Accurate ✓

3. Data Wrangling

Data Wrangling

Data wrangling is the process of cleaning, restructuring and enriching available raw data into a format that is more usable. This allows users to ease and quicken their decision-making process and get better data insights.

Practitioners consistently report that around 80% of the effort involved in dealing with data is cleaning it up, massaging it and getting it ready for use.

The Data Wrangling Steps

- **Discovering**
- Structuring
- Cleaning
- Enriching
- Validating
- Publishing

Allows data analyst to understand the data and how useful it is for analytic exploration and analysis.

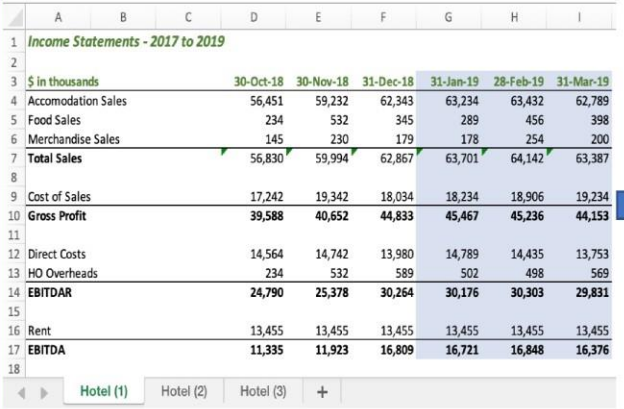
This is where a deeper understanding of the data is obtained.

The Data Wrangling Steps

- Discovering
- **Structuring**
- Cleaning
- Enriching
- Validating
- Publishing

Gives the ability to format data of all shapes and sizes to work with traditional applications. In most cases, raw data is obtained in a haphazard manner with no structure. It may come in a form that may look great to human eyes, but is a disaster for analytical softwares such as Excel and Alteryx. Structuring data may include splitting columns, pivoting rows and deleting fields, among other things.

Below is an example of transforming a typical excel income statement schedule to a “flat file” format to be imported into an analytical tool.



	A	B	C	D	E	F	G	H	I
1	*Income Statements - 2017 to 2019*								
2									
3	$ in thousands			30-Oct-18	30-Nov-18	31-Dec-18	31-Jan-19	28-Feb-19	31-Mar-19
4	Accomodation Sales			56,451	59,232	62,343	63,234	63,432	62,789
5	Food Sales			234	532	345	289	456	398
6	Merchandise Sales			145	230	179	178	254	200
7	**Total Sales**			56,830	59,994	62,867	63,701	64,142	63,387
8									
9	Cost of Sales			17,242	19,342	18,034	18,234	18,906	19,234
10	**Gross Profit**			**39,588**	**40,652**	**44,833**	**45,467**	**45,236**	**44,153**
11									
12	Direct Costs			14,564	14,742	13,980	14,789	14,435	13,753
13	HO Overheads			234	532	589	502	498	569
14	**EBITDAR**			**24,790**	**25,378**	**30,264**	**30,176**	**30,303**	**29,831**
15									
16	Rent			13,455	13,455	13,455	13,455	13,455	13,455
17	**EBITDA**			**11,335**	**11,923**	**16,809**	**16,721**	**16,848**	**16,376**
18									

Hotel (1) | Hotel (2) | Hotel (3) | +

	A	B	C	D
1	Hotel	Line Item	Date	Value
2	Hotel (1)	Accomodation Sales	30/09/2019	10942.32
3	Hotel (1)	Accomodation Sales	25/07/2019	13242.21
4	Hotel (1)	Accomodation Sales	21/03/2019	1238
5	Hotel (1)	Accomodation Sales	05/03/2019	56321.21
6	Hotel (1)	Accomodation Sales	06/03/2019	2314.1
7	Hotel (1)	Accomodation Sales	25/07/2019	12312.12
8	Hotel (1)	Accomodation Sales	20/09/2018	34341
9	Hotel (1)	Accomodation Sales	13/11/2018	4312.34
10	Hotel (1)	Accomodation Sales	12/05/2019	312312.43
11	Hotel (1)	Accomodation Sales	13/08/2018	43578.3
12	Hotel (1)	Accomodation Sales	02/02/2018	3432657.87
13	Hotel (1)	Accomodation Sales	05/03/2019	434.3
14	Hotel (1)	Accomodation Sales	06/03/2019	3454.65
15	Hotel (1)	Accomodation Sales	09/12/2017	3431.1
16	Hotel (1)	Accomodation Sales	06/06/2018	76534.21
17	Hotel (1)	Accomodation Sales	19/07/2018	3331.1

The Data Wrangling Steps

- Discovering
- Structuring
- **Cleaning**
- Enriching
- Validating
- Publishing

Involves identifying data quality issues and removing any outliers in the dataset which can skew and distort the results of the analysis. In addition, any null values will have to be changed, and the formatting will have to be standardized in order to improve data quality.

A data set is said to be clean if:

1. Each piece of the data is in a format that a software can 'read'

Common formatting errors:

- A zero is reported as "NULL" or a blank cell
- A number is entered as a text object
- A piece of data has an extra character
- A "leading space" – a blank space before the piece of data

2. There is 100% consistency in the format of the data items in each record

Common inconsistencies:

- The *Date* field has dates in different formats
- The data is in different units (for example, a mix of centimeters and inches)

3. It is complete and accurate

Common incompleteness & inaccuracy issues:

- One-off errors – for example, missing data in a cell or missing block of rows
- A pattern – for example, every third row is missing

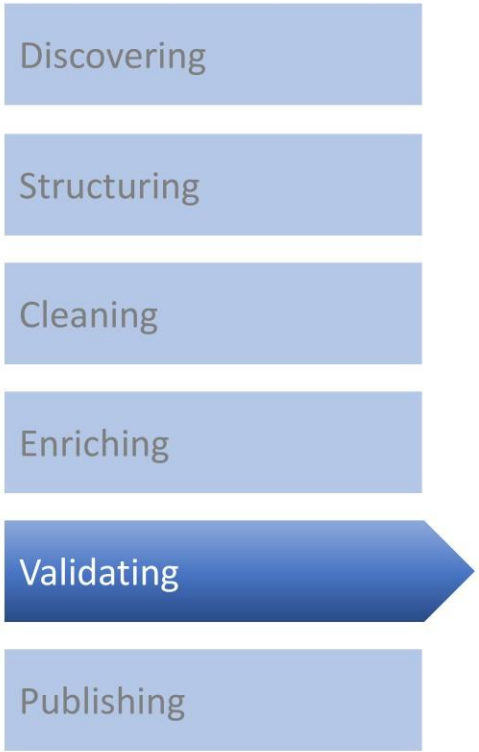
The Data Wrangling Steps

- Discovering
- Structuring
- Cleaning
- **Enriching**
- Validating
- Publishing

Once data has been cleaned, it will have to be enriched.

This entails taking stock of what is in the data and strategizing to determine whether it must be augmented using some additional data in order to make it better. The existing dataset may be enriched by joining and aggregating multiple data sources.

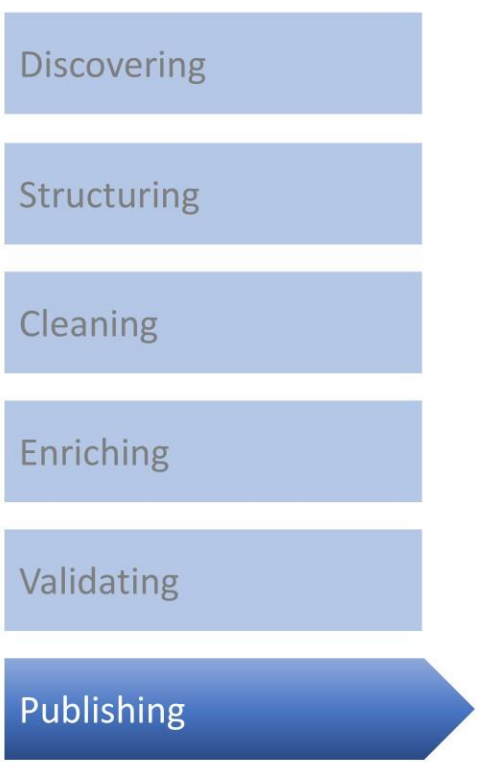
The Data Wrangling Steps



It is essential to manually check datasets and perform a final human check for any missing or mismatched data that wasn't corrected during the transformation process.

Validating that the output dataset has the intended structure, consistency and content before publishing is a key step in the data wrangling process.

The Data Wrangling Steps



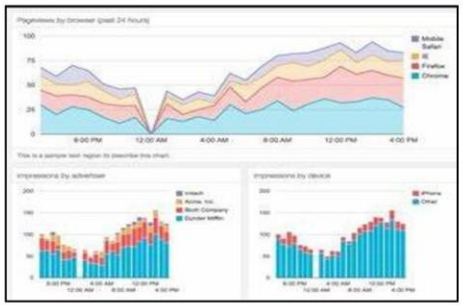
Once the data has been successfully structured, cleaned, enriched and validated, it is time to publish the wrangled output for downstream analysis.

The data wrangling process allows users to see a wide range of statistics and analytics that aim to add value to the organization and make better decisions.

Meet the Data Analyst

aka Data Wrangler

Data wrangling is the process of transforming raw data into information that is usable for informed decision making.



Raw Data
(Structured &
Unstructured)

Data Analyst
wrangling with
Raw Data

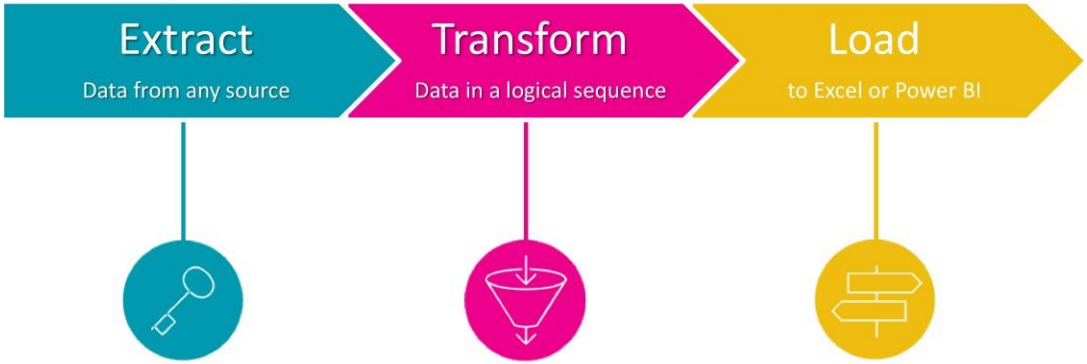
Information Output
after
A LOT of work

4. Power Query

Power Query

to the rescue

Power Query is a strong ETL tool and the modern Accountant's best friend.



What is Power Query?

Power Query is a business intelligence tool available in Excel that allows you to import data from many different sources and then clean, transform and reshape your data as needed.

It allows you to set up a query once and then reuse it with a simple refresh. It is pretty powerful.

Power Query can import and clean millions of rows of data for later analysis and presentation.

The user interface is intuitive and well laid out so it's really easy to learn. It has an incredibly short learning curve when compared to other Excel tools like some of the more complex formulas in Excel or VBA.

The best part about it, is you don't need to learn or use any code to do any of it. The power query editor records all your data transformations step by step and converts them into its code language, called 'M'. This is very similar to how the Macro recorder works in Excel to generate VBA code.

If you want to edit or write your own M code, you certainly can, but you don't need to worry about this as a beginner and certainly not for your exam.

What can Power Query do?

```
Sales ID,Product ID,Date,Comm Rate,Sales
10001,10030001,2016-08-28,25.0%,65.5
10002,10030002,2017-11-01,25.0%,59.5
10003,10030001,2016-01-29,25.0%,65.5
10004,10010001,2016-03-12,10.0%,1499.5
10005,10030003,2016-03-04,25.0%,16.5
10006,10030004,2017-09-30,25.0%,56.5
10007,10030005,2016-12-29,25.0%,9.6
10008,10030001,2017-02-01,25.0%,65.5
10005,10030005,2016-09-06,25.0%,9.6
10004,10030006,2016-05-20,25.0%,12.9
```

Imagine you get a sales report uploaded in a CSV file format on a monthly basis that looks like the screenshot to the left. Based on data in this file, you are asked to calculate commission due to each salesperson and present it as below. If doing this in Excel …

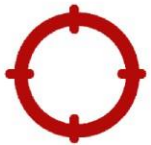
Sales Person	Total Commission
Ryan Bohan	$1,694.23
Arron Mattin	$1,139.43
Collin Abthorpe	$411.15
Doug Howis	$2,779.53
Max Renton	$2,999.73
Glen Thomke	$1,939.93
Reilly Wynne	$668.75
Harvey Caven	$644.18
Raquel Lilywhite	$2,315.28
Johanna Marten	$1,101.20
Isaac Tillard	$485.65

… each month you'd have to go to the folder where the file is uploaded, open the file and copy its contents into Excel. You would then use the text-to-column feature in Excel to split out the data into new columns. The sales report only shows [Sales ID], so you would need to add a new column to the Excel file and use a VLOOKUP to get the salesperson associated with each [Sales ID]. You would then summarize the total sales by each salesperson and calculate the commission to pay out.

Maybe it only takes an hour or so each month, but it's pretty mindless work that's boring, unenjoyable and takes away from time you can actually spend analyzing the data and producing meaningful insight.

What can Power Query do?

With Power Query, all of the steps mentioned on the previous page can be automated down to a click of the **Refresh** button on a monthly basis.



Define once

Consume anytime

All you need to do is build the query once and reuse it each month, thereby saving an hour of time each and every month.

```
Sales ID,Product ID,Date,Comm Rate,Sales
10001,10030001,2016-08-28,25.0%,65.5
10002,10030002,2017-11-01,25.0%,59.5
10003,10030001,2016-01-29,25.0%,65.5
10004,10010001,2016-03-12,10.0%,1499.5
10005,10030003,2016-03-04,25.0%,16.5
10006,10030004,2017-09-30,25.0%,56.5
10007,10030005,2016-12-29,25.0%,9.6
10008,10030001,2017-02-01,25.0%,65.5
10005,10030005,2016-09-06,25.0%,9.6
10004,10030006,2016-05-20,25.0%,12.9
```

Go from raw data to cleaned and summarised with one click

Sales Person	Total Commission
Ryan Bohan	$1,694.23
Arron Mattin	$1,139.43
Collin Abthorpe	$411.15
Doug Howis	$2,779.53
Max Renton	$2,999.73
Glen Thomke	$1,939.93
Reilly Wynne	$668.75
Harvey Caven	$644.18
Raquel Lilywhite	$2,315.28
Johanna Marten	$1,101.20
Isaac Tillard	$485.65

Power Query

what makes it stand out?

Simple
Intuitive and easy to use

Practical
Created for the business user having no coding skills

Compatible
Same code works everywhere

Rich
New features and data connectors are constantly being added

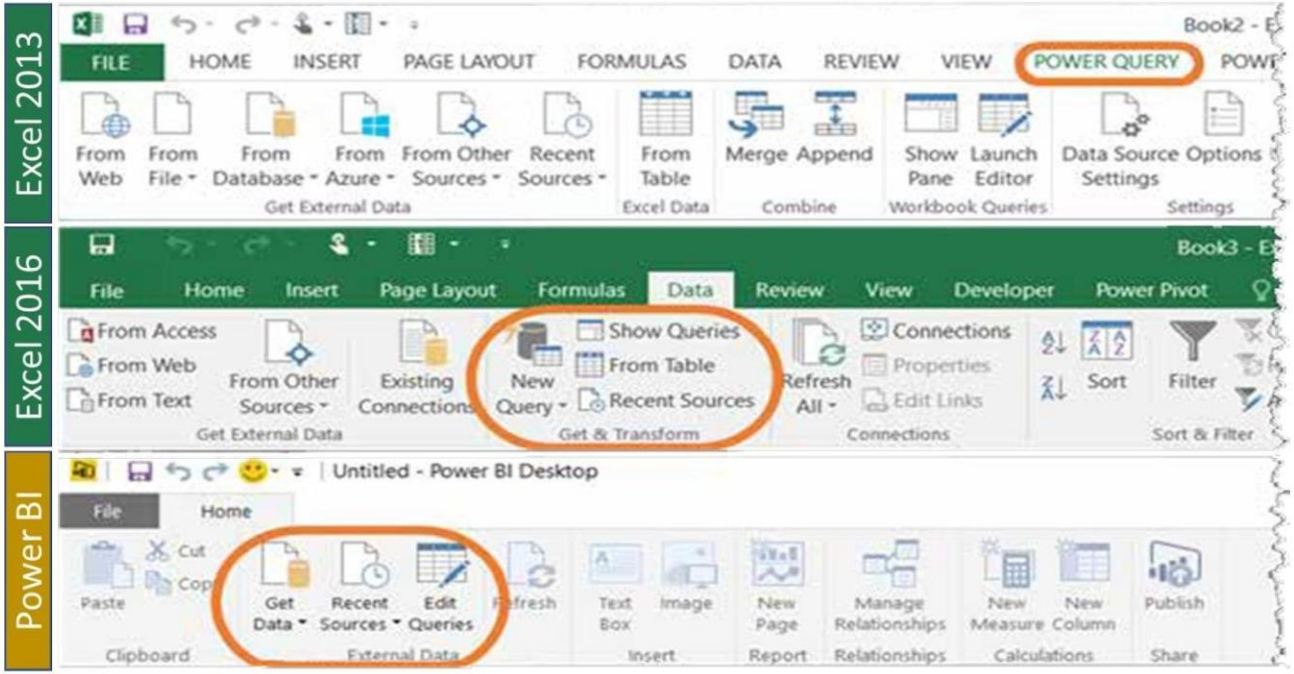
Familiar
Easy to grasp features and learn the programming language (called M)

Scalable
Same solution can work on the desktop, on the web or on mobile media

Power Query

where can you find it?

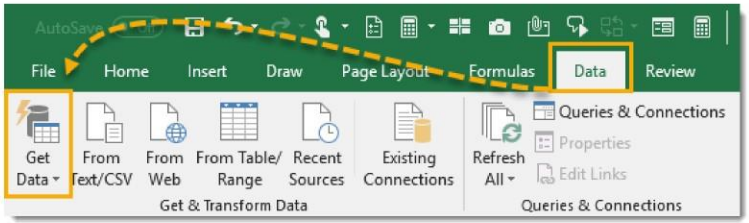
Power Query was initially released after Excel 2013 as a free add-in for Excel, and it was (rather shockingly) backward compatible with Excel 2010. In both of those versions, a unique ribbon tab holds all the Power Query commands. Because Power Query is so incredibly useful, it only made sense to integrate the tool into Excel 2016. Due to concerns about bloating the user interface, though, Power Query was not given its own ribbon tab in Excel 2016 but was instead squished down into the Get & Transform group on the Data tab. The image above shows where Power Query is accessible in both Excel & Power BI.



Importing Data with Power Query

Importing your data with Power Query is simple.

Excel provides several different data connections that are accessible from the **Data** tab and can be found under the **Get Data** command. Note that the available data connection options for Get Data is dependent on your version of Excel.



In general, you can …

… Get data from **a single file** such as an Excel workbook, Text or CSV file, XML and JSON files. You can also import multiple files from within **a single folder**.

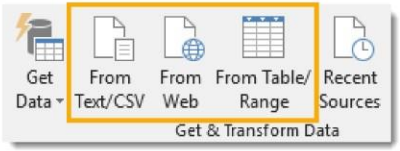
… Get data from **various databases** such as SQL Server, Microsoft Access, Analysis Services, SQL Server Analysis Server, Oracle, IBM DB2, MySQL, PostgreSQL, Sybase, Teradata and SAP HANA databases.

… Get data from **Microsoft Azure** and from other **online services** like Sharepoint, Microsoft Exchange, Dynamics 365, Facebook and Salesforce.

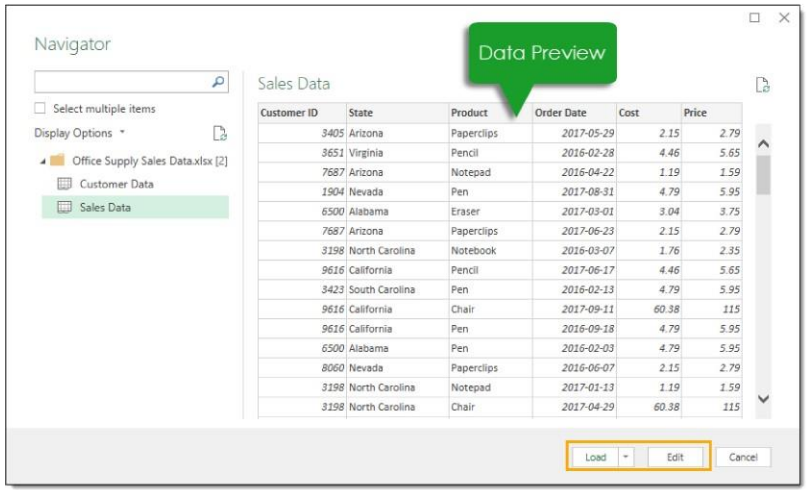
… Get data from **other sources** like a table or range inside the current workbook, from the web, a Microsoft Query, Hadoop, OData feed, ODBC and OLEDB.

Importing Data with Power Query

There are some of the more common data connection options available within top level of the Ribbon, as found in **Get & Transform Data** section of the Data tab.



From here, we can easily access **From Text/CSV**, **From Web** & **From Table/Range** queries. These are duplicated outside of the Get Data command simply for the sake of convenience.

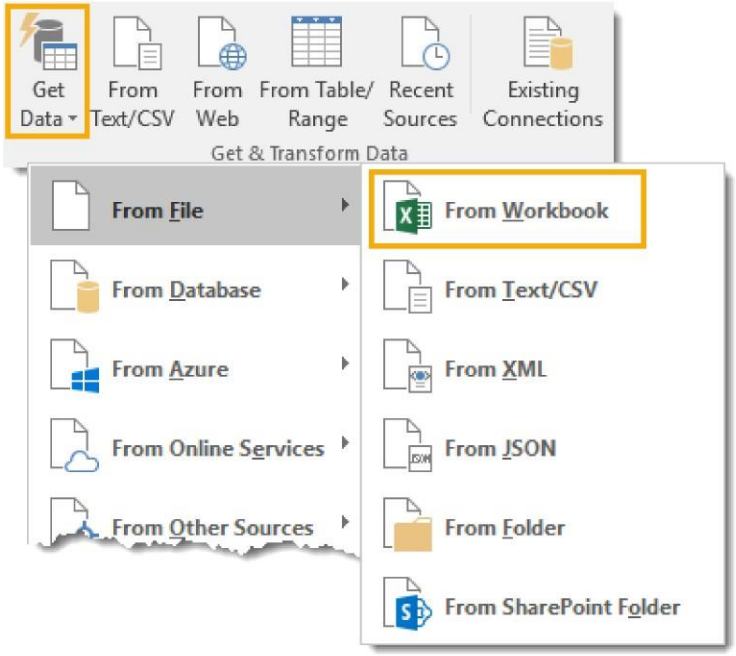


Depending on which type of data connection you choose, Power Query will guide you through the data connection setup; and there might be several options to select during the process.

At the end of the setup process, you will come to the **Data Preview** window. You can preview a sample of the data here to confirm it is what you're expecting. You can then load the data as-is by pressing the **Load** button, or you can proceed to the query editor to apply any data transformation steps by pressing the **Edit** button.

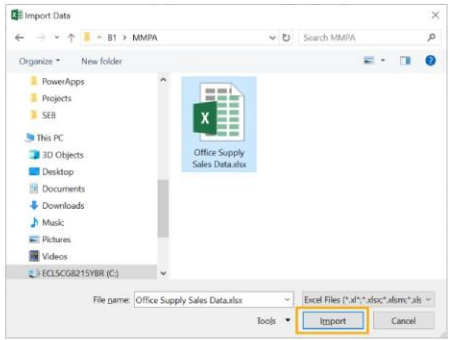
Importing Data in an Excel File

Let's take a look at importing some data from an Excel workbook in action.



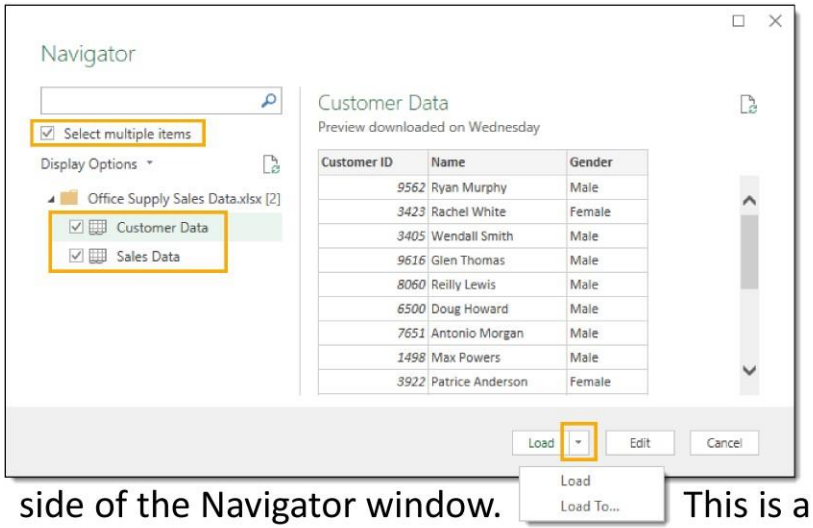
We shall import an Excel file called Office Supply Sales Data.xlsx. It contains sales data on one sheet called Sales Data and customer data on another sheet called Customer Data. Both sheets of data start in cell A1 and the first row of the data contains column headers.

Go to the **Data** tab and select the **Get Data** command in the Get & Transform Data section of the Ribbon. Then go to **From File** and choose **From Workbook**.



This will open a **Import Data** dialog form where you can navigate to the file you want to import. Select the file and click on the **Import** button.

Importing Data in an Excel File



After selecting the file you want to import, the data preview **Navigator** window will open. This will give you a list of all the objects (data tables and ranges) available to import from the workbook.

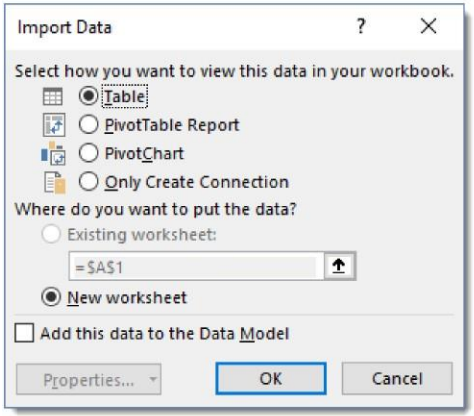
Check the box to **Select multiple items** since we will be importing data from two different worksheets. Now we can check both the Customer Data and the Sales Data.

When you click on either of the objects in the workbook, you can see a preview of the data for it on the right hand side of the Navigator window. This is a great way to perform a reasonability check to validate whether or not you've got the correct file.

When you're satisfied that you've got everything you need from the workbook, you can press either one of the **Edit** or **Load** buttons. The Edit button will take you to the Query Editor where you can transform your data before loading it. Pressing the Load button will load the data into tables in new sheets in the workbook.

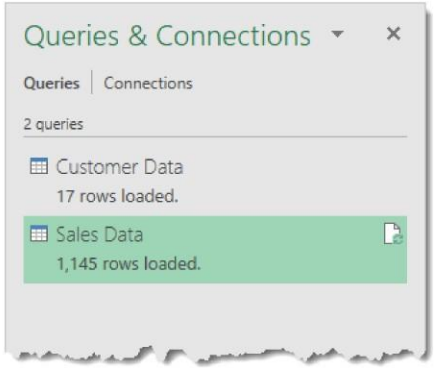
Importing Data in an Excel File

For now, let's bypass the Query Editor and go straight to loading the data into Excel.



Press the small arrow next to the Load button to access the **Load To…** options. This will present you with a **Import Data** form with a few more loading options. We will choose to load the data into a **Table** in a **New worksheet**. However, note there are several other options to choose from.

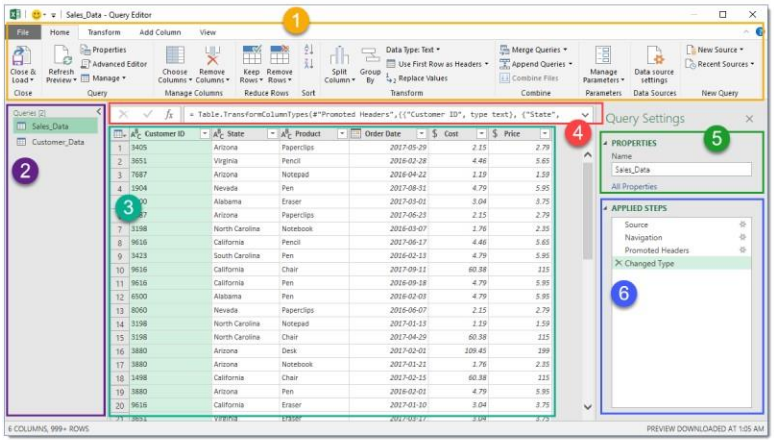
As seen in the screenshot to the left, you may also load the data directly into a **PivotTable Report**, or a **PivotChart**, or you can completely avoid loading the data and just **Only Create Connection** to the data.



As we click on OK, the two tables are loaded into two new worksheets in Excel and we also have two queries which can quickly be refreshed if the data in the original workbook is ever updated. Excel shows us the queries in a side panel called **Queries & Connections**. We can later select any one of these queries and click on **Edit** to go back and make changes to the query as required.

The Query Editor

After connecting to your data and selecting the Edit option, you are presented with the **Query Editor**. This is where any data transformation steps will be created or edited.



There are 6 panes in the Editor you need to be familiar with:

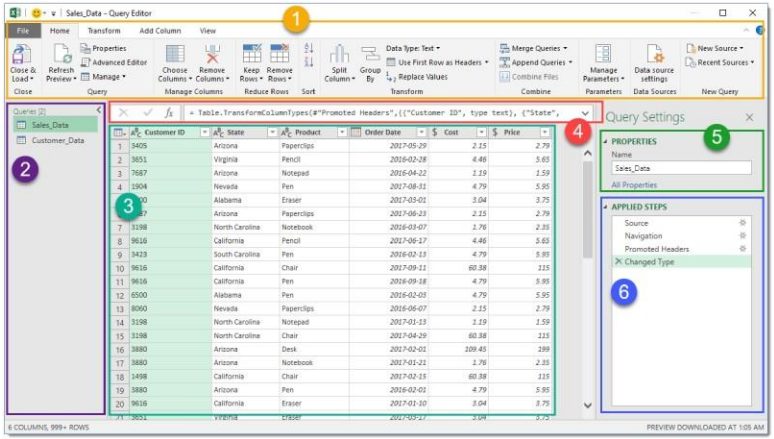
1. Ribbon – The user interface for the Editor is quite similar to Excel and uses a visual ribbon style command center. It organizes data transformation commands and other Power Query options in the remaining 5 panes.

2. Query List – This pane lists all the queries in the current workbook. You can navigate to any query from this pane to begin editing it. You can collapse and expand this pane.

3. Data Preview – This pane is where you see a preview of the data with all the transformation steps currently applied. You can also access a lot of the transformation commands here either from the filter icons in the column headings or with a right click on the column heading.

The Query Editor

We explored 3 of the 6 panes in **Query Editor** on the previous page. Below are the remaining 3 …



4. Formula Bar – This is where you can see and edit the M code of the current transformation step. Each transformation you make on your data is recorded and appears as a step in the Applied Steps pane.

5. Properties – This is where you can name your query. When you close and load the query to an Excel table, Power Query will create a table with the same name as its source query if the table name isn't already taken. Giving a meaningful name to the query is important as this name is what the M code will refer to if we were to query it in another query.

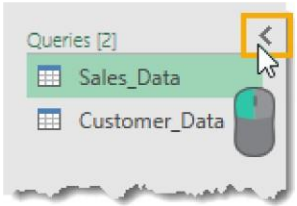
6. Applied Steps – This pane shows a sequential list of all the transformation steps that have been applied to the data. You can move through the steps here and view the changes in the Data Preview pane. You can also delete, modify or reorder any steps in the query here.

The Query List

The Query List has abilities other than just listing out all queries in the current workbook.

One of the primary functions of the Query List is Navigation. There's no need to exit the Query Editor to switch which query you're working on. You can click on any query in the Query List to switch to it. The query you're currently on is highlighted in light green color.

When you exit the Query Editor, all changes made in all the queries you edited will automatically be saved.

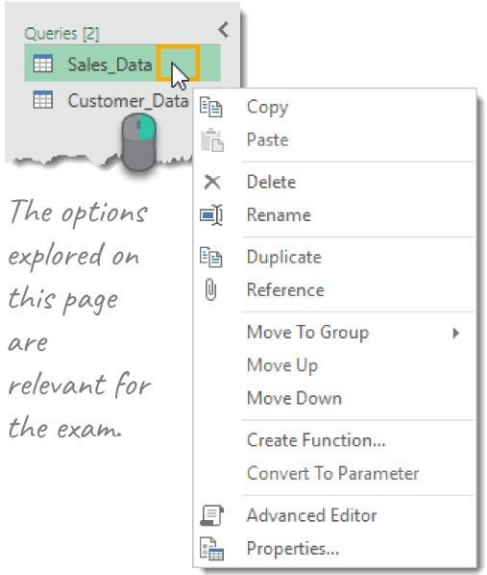


You can hide the Query List to create more room for the Data Preview.

To do so, simply click on the small arrow in the upper right corner of the Query List to toggle the list between hidden (collapsed) and visible (expanded).

The Query List

If you right click on any query in the Query List pane, there are a variety of options available...



The options explored on this page are relevant for the exam.

Copy and **Paste** – Copy and paste a query to make another copy of it.

Delete – Delete the query. Note that if you accidentally delete a query, there's no undo button for it. However, you can exit the Query Editor without saving to restore your query.

Rename – Rename your query. This is the same as renaming it within the Properties area of the Query Editor.

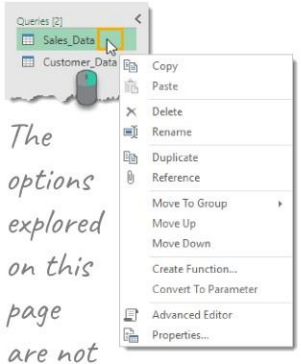
Duplicate – Make another copy of the query. This is the same as copy and paste but completes the process in one step.

Move To Group – Place your queries into a folder like structure to keep them organized when the Query List gets large.

Move Up and **Move Down** – Rearrange the order your queries appear in the Query List or within the folder groups. This can also be done by dragging and dropping the query to its new location.

The Query List

If you right click on any query in the Query List pane, there are a variety of options available...



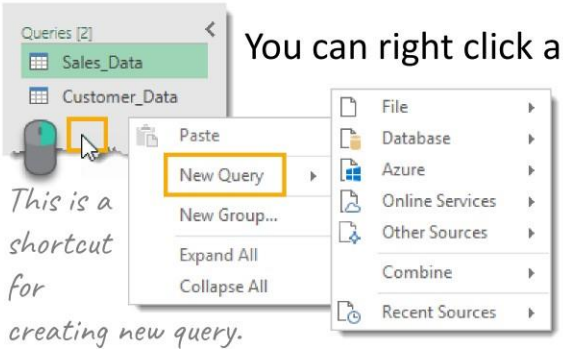
The options explored on this page are not relevant for the exam.

Create Function… – Turn your query into a query function. This allows you to pass a parameter to the query and return results based on the parameter passed.

Convert To Parameter – Allows you to convert queries to parameters.

Advanced Editor – Open the advanced editor to edit the M code for the query.

Properties… – Allows you to change the query name, add a description text and enable Fast Data Load option for the query.



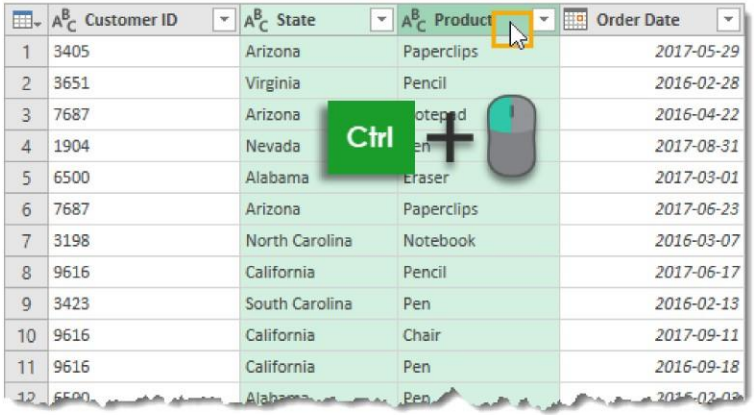
This is a shortcut for creating new query.

You can right click any empty area in the Query List to create a new query.

The menu options that are made available to you after clicking on New Query are similar to what you would get if you were to click on the Get Data command in the Get & Transform Data section of the Ribbon.

The Data Preview

The main job of the Data Preview pane is to apply transformation steps to your data and show a preview of each step you're applying.



	Customer ID	State	Product	Order Date
1	3405	Arizona	Paperclips	2017-05-29
2	3651	Virginia	Pencil	2016-02-28
3	7687	Arizona	otepad	2016-04-22
4	1904	Nevada	en	2017-08-31
5	6500	Alabama	Eraser	2017-03-01
6	7687	Arizona	Paperclips	2017-06-23
7	3198	North Carolina	Notebook	2016-03-07
8	9616	California	Pencil	2017-06-17
9	3423	South Carolina	Pen	2016-02-13
10	9616	California	Chair	2017-09-11
11	9616	California	Pen	2016-09-18
12	6500	Alabama	Pen	2016-02-03

In the data preview area, you can select columns using a few different methods. Each column will be highlighted in a light green color when it is selected.

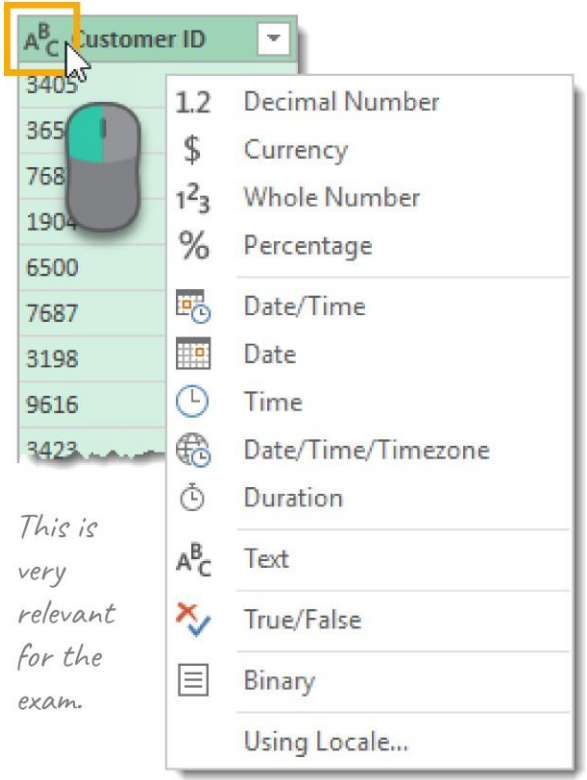
Select a single column with a left click on the column heading.

Select multiple adjacent columns with a left click on the first column heading, then hold Shift and left click on the last column heading.

Select multiple non-adjacent columns by holding Ctrl + left click on any column headings you want to select.

You can then apply any relevant data transformation steps on selected columns from the Ribbon. Certain commands can also be accessed with a right click on the column heading. Commands that are not applicable to your selected column or columns will appear grayed out in the Ribbon.

The Data Preview



This is very relevant for the exam.

Each column in the Data Preview pane has a data type icon to the left of the column heading. You can left click on this icon to change the data type of the column.

You can choose from any of the data types listed in the screenshot to the left.

A word on: Using Locale…

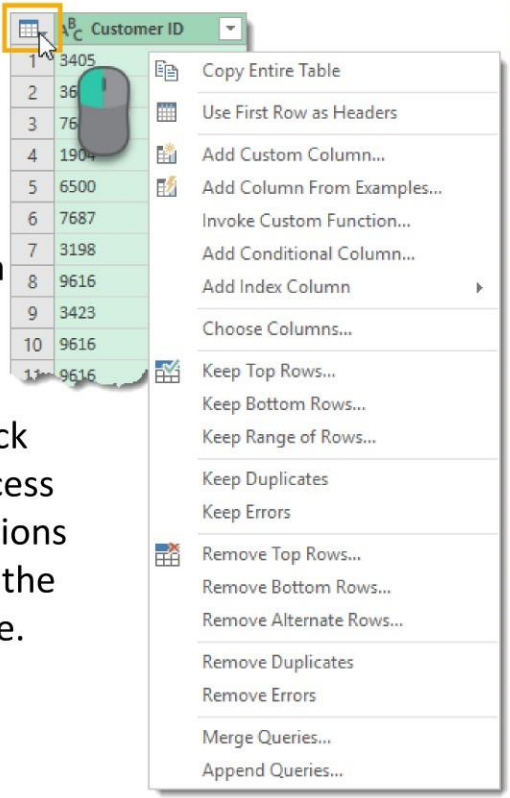
Selecting the 'Using Locale…' option allows you to set the data type format using the specific convention for different locations. You can use locale for any data type.

For example, if you wanted to display the date in the American locale's M/D/YYYY format then you would select 'English (United States)' as the locale.

The Data Preview

There's a small 'table' icon in the top left corner of the Data Preview pane.

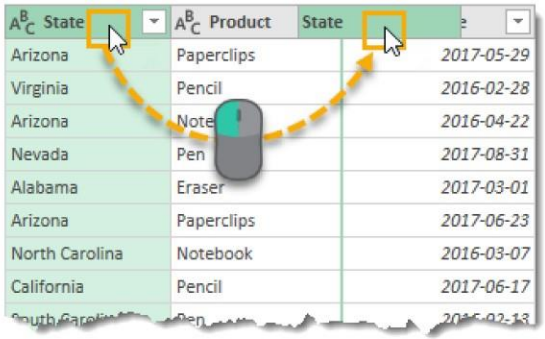
You can click on it to access various actions that affect the whole table.



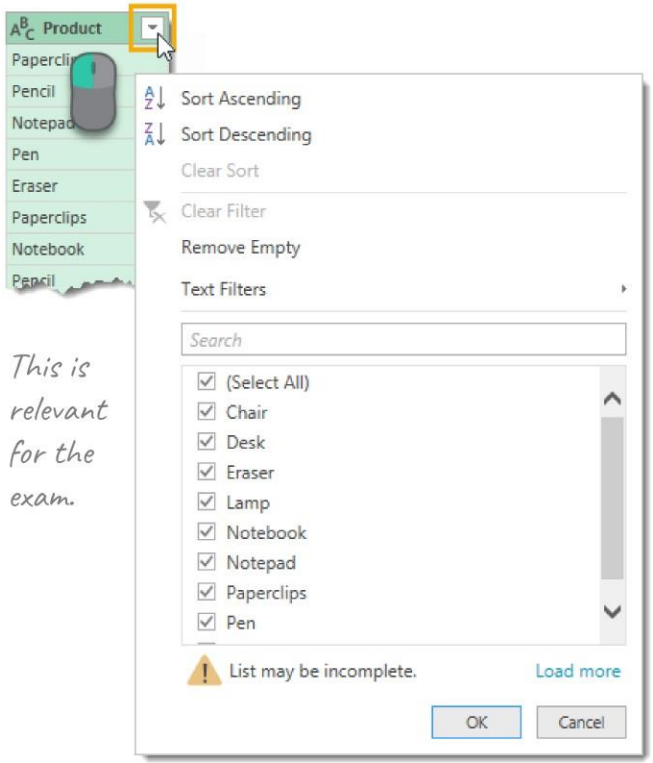
Renaming any column heading is really easy.

Double click on any column heading; then type a new name for the column heading and press Enter when done.

You can change the order of columns with a left click and drag-and-drop action. The green border between two columns will become the new location of the dragged column when you release the left click.



The Data Preview



This is relevant for the exam.

Each column has a filter toggle to the right of the column heading, that looks like a drop-down arrow. Click on it to sort and filter your data.

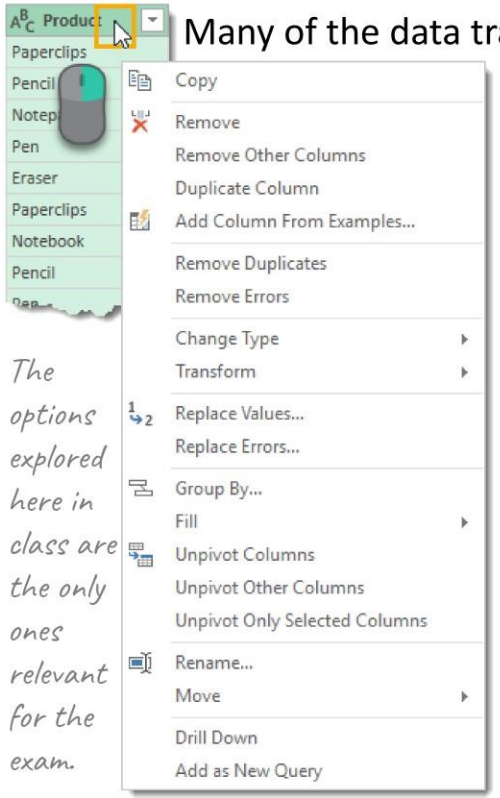
The filter options is very similar to the filter options found in a regular Excel spreadsheet. It also works exactly the same way.

Note:

The list of items shown on initial drop-down is always based on a sample of the data displayed - so it may not always contain all available items in the data population.

You can load more data by clicking on the **Load more** text in blue.

The Data Preview



The options explored here in class are the only ones relevant for the exam.

Many of the data transformations found in the Ribbon menu are also accessible from the Data Preview pane, using a right click on the column heading. Following are some of the available options that we shall explore in class:

Remove – Remove the selected column(s).

Remove Other Columns – Remove all columns other than selected column(s).

Duplicate Column – Copy and paste the column in one step.

Remove Duplicates – Remove rows based on duplicate values in selected column.

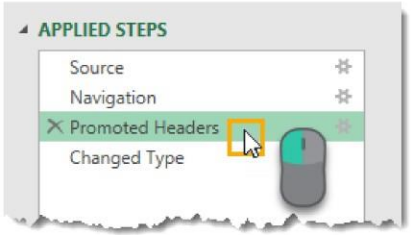
Remove Errors – Remove rows based on error values in selected column.

Replace Values… – Find and Replace defined value in column with provided value.

Replace Errors… – Find and Replace error values in column with provided value.

Fill – Fill empty cells in a column, either up or down, with the data in preceding non-empty cell.

The Applied Steps



Any transformation you make to your data will appear as a step in the Applied Steps pane. It also allows you to navigate step-by-step through your query.

Click on any step in the Applied Steps pane, and the Data Preview pane will update to show all data transformations up to and including that step.

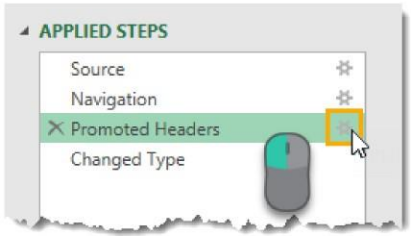
You can insert new steps into the query at any point by selecting the previous step and then creating the transformation in the data preview. Power Query will then ask if you want to insert this new step. Careful though, as this may break the following steps in the Applied Steps pane that refer to something you changed.

You can delete any steps that were applied using the ✕ on the left hand side of the step name in the Applied Steps pane.

You must exercise caution when deleting a step, since if any of the following steps depend on the step you're trying to delete, you will break your query.

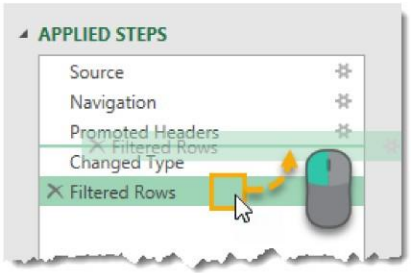
This is where Delete Until End from the right click menu can come in handy.

The Applied Steps



A lot of transformation steps available in power query will have various user input parameters and other setting associated with them. If you apply a filter on the product column to 'show all items not starting with Pen', you might later decide you need to change this filter step to 'show all items not equal to Pen'. You can easily make all such edits from the Applied Steps pane.

Some of the steps will have a small gear icon on the right hand side. This allows you to edit and change the inputs and settings of that step at any point in time.

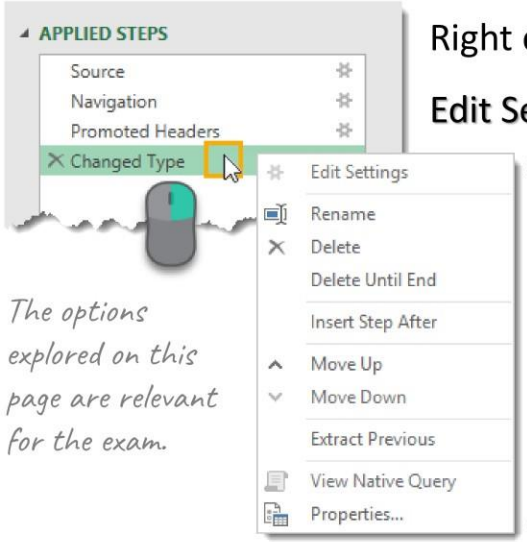


You can also rearrange the sequential order in which the steps are performed in your query.

Just left click on any step and drag it to a new location. A green line will appear between steps to indicate the new location where it is being dragged to.

This is another example where you must exercise caution, since a lot of following steps depend on previous steps, and because of this changing the sequential order of steps can create errors.

The Applied Steps



The options explored on this page are relevant for the exam.

Right click on any step to access a menu of options.

Edit Settings – This allows you to edit the settings of the step similar to using the gear icon on the right hand side of the step.

Rename – This allows you to rename the step's label to something meaningful, so you can easily identify what the step is doing.

Delete – This deletes the current step.

Delete Until End – This allows you to delete the current step plus all steps up until the end. Since steps can depend on previous steps, deleting all steps after a step is a good way to avoid any errors.

Insert Step After – This allows you to insert a new step after the current step in the query.

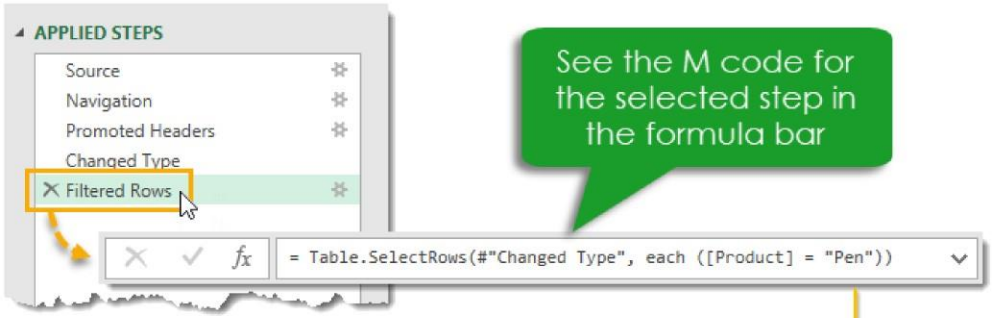
Move Up and **Move Down** – This allows you to rearrange the sequential order of steps in the query.

Extract Previous – This allows you to create a new copy of the query up to the selected step. It can be a really useful option when you want to create a new query based on the logic followed in an existing one.

The Formula Bar

When you click on different steps of the transformation process in the Applied Steps pane, the **Formula Bar** updates to show the M code that was created for that step.

If the M code generated is longer than the formula bar, you can expand the formula bar using the arrow toggle on the right hand side.

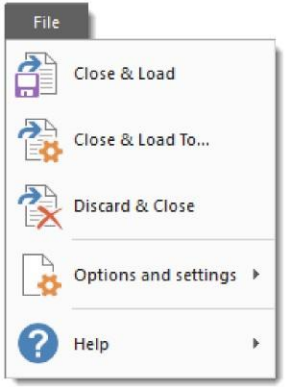


You can edit the M code for a step directly from the formula bar without the need to open the advanced editor.

In our example above , we've changed our filter from 'Pen' to 'Chair' by typing in the formula bar and then pressing Enter or using the ✓ on the left to confirm the change.

Press Esc or use the ✕ on the left to discard any changes.

The File tab



The File tab contains various options for saving any changes made to your queries as well as Power Query options and settings.

Close & Load – This will save your queries and load your current query into an Excel table in the workbook.

Close & Load To… – This will open the Import Data menu with various data loading options to choose from.

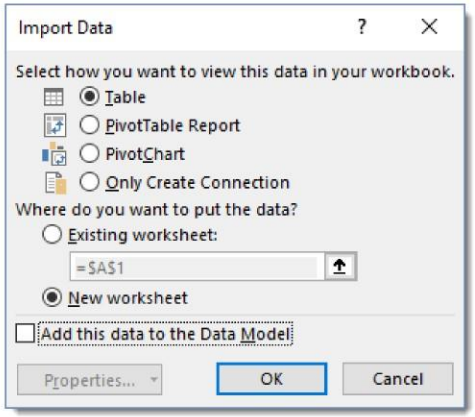
Discard & Close – This will discard any changes you made to the queries during your session in the editor and close the editor.

Note:

- Regardless of saving the query in Power Query, you will still need to save the Excel workbook in the regular way to keep any changes made to the queries once you close the Excel workbook.
- **Close & Load** and **Close & Load To…** commands are also available from the **Home** tab.

Data Loading Options

When you use the Close & Load To option to exit the editor, this will open the Import Data menu.



You can choose to load the query to a Table, PivotTable Report, PivotChart or Only Create Connection for the query. The connection only option means that there is no data output to the workbook, but you can still use this query in other queries. This is a good option if the query is an intermediate step in a data transformation process.

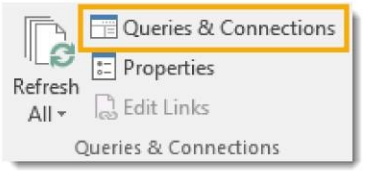
You are able to select the location to load to in your workbook if you select an option other than connection only. You can choose a cell in an Existing worksheet or load it to a New worksheet that Excel will create automatically.

The other option you get is the Add this data to the Data Model. This will allow you to use the data output in Power Pivot and use other Data Model functionality like building relationships between tables.

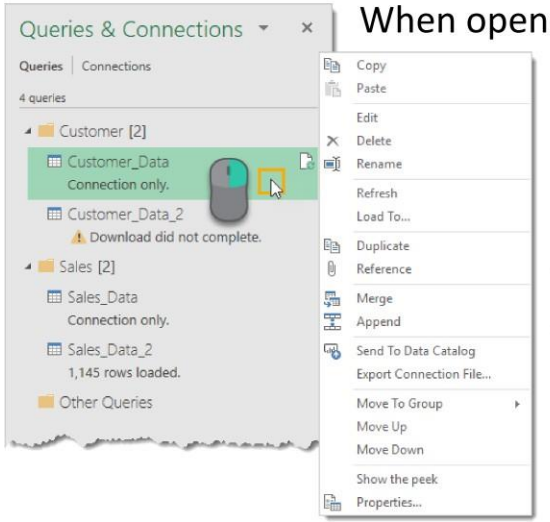
The Data Model is Excel's new and efficient way of storing and using large amounts of data. However, this is beyond our scope of discussion.

Queries & Connections

When you're working outside of the power query editor, you can see and interact with all the queries in the workbook through the Queries & Connections window.



To open this, go to the Data tab in the regular Excel ribbon, then press the Queries & Connections command button found in the Queries & Connections section.

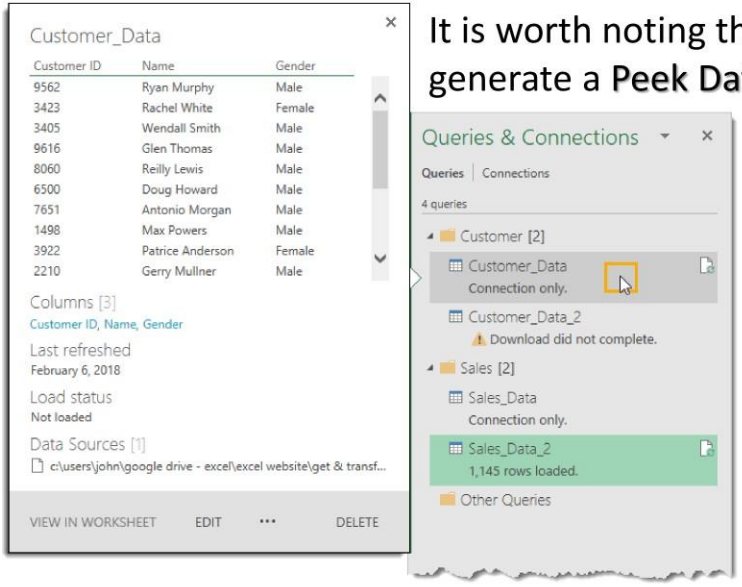


When opened the Queries & Connections window will be docked to the right hand side of the workbook. You can undock it by left clicking on the title and dragging it. You can drag it to the left hand side and dock it there or leave it floating. You can also resize the window by left clicking and dragging the edges.

The Queries & Connections window is very similar to the query list in the editor and you can perform a lot of the same actions with a right click on any query.

One point worth noting is that the Load To option is not available in the right click menu.

Queries & Connections



It is worth noting that when you hover over a query with your mouse, Excel will generate a **Peek Data Preview**, plus show some basic information about the query.

Data Preview – This is a live preview of the data similar to when first setting up a query.

Columns – This will give you a list of all the columns contained in the final results of the query along with a count of how many columns there are. Clicking on any of them will highlight the column in the data preview.

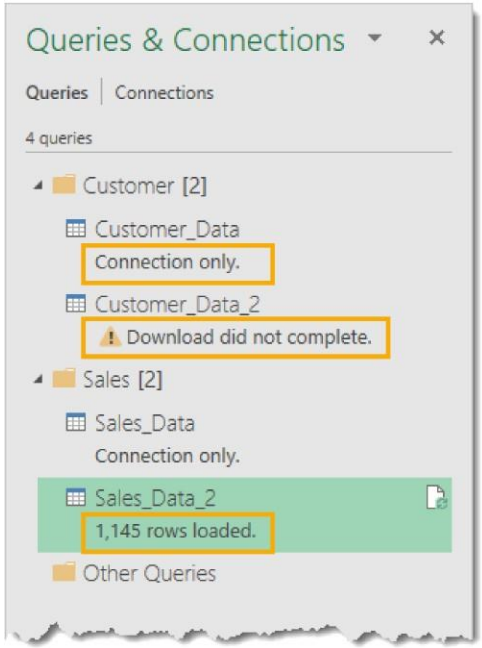
Last Refreshed – This will tell you when the data was last refreshed.

Load Status – This displays whether the data is loaded to a table, pivot table, pivot chart or is connection only.

Data Sources – This will show you the source of the data along with a total count of files.

View in Worksheet – This will take you to the worksheet if query is loaded to a Table, PivotTable or PivotChart.

Queries & Connections



There are also some useful messages displayed in the Queries & Connections window for each query.

It will show you:

- if the query is a connection only query,
- if there were any errors when the query last ran, or
- how many rows loaded if it is not a connection only query.

Power Query
how does it work?

Basic Approach:

- Looks at a preview of your data
- Imports and Transforms the data
- Applies “Types” to your Data
- Loads the query to connection, data model or (in Excel) a table

To Refresh a Power Query:

- Click the Refresh button
- Power Query runs the updated source file through the same steps!

Power Query
what shall we learn?

- Loading Data
- Splitting Data
- Reshaping Data
- Combining Data
- Appending Data

Loading Data

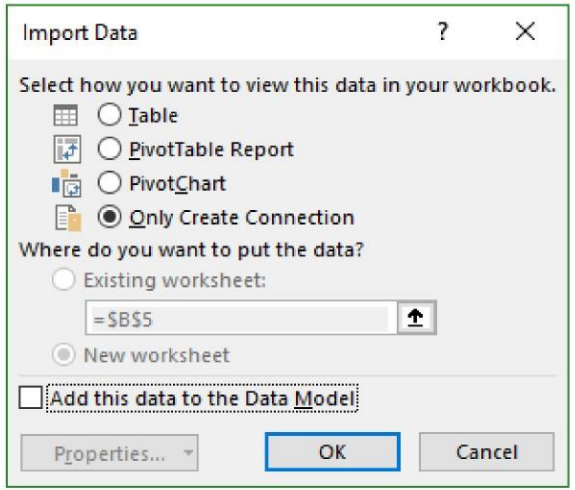
Creating a Staging Query

Step	Tasks to perform
1	Prepare the Source Data

- Connect to your Data source
- Perform any desired transformations
- Name your query

2	Load as Staging query

- Go to Home → Close & Load → Close & Load To…
- Choose: Only Create Connection
- Uncheck: Add this data to the Data Model



Note: Staging queries …
- do not consume any processor or RAM until they are called by another query.
- can be very helpful in order to act as a starting point for other queries.

Loading Data

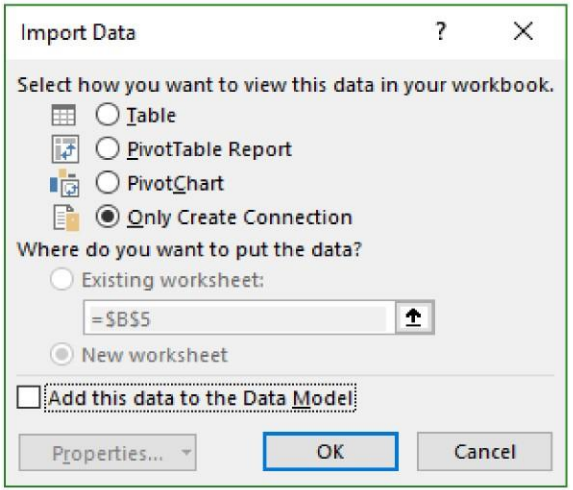
Modifying Query Load Destinations

Step	Tasks to perform
1	Display the Queries & Connections Pane

- Excel 2010/2013 Power Query →Show Pane
- Excel 2016 Data →Show Queries
- Excel 2019/365 Data →Queries & Connections

2 Modify the Load Destination

- Right click the desired query →Load To…
- Modify the load behavior as desired



Note: If you accidentally loaded a query to a table, you can convert it to a Connection Only query by simply deleting the worksheet.

Splitting Data

Splitting Records into Columns

Step	Tasks to perform
1	Right click the Values column
2	Choose to Split Column →by Delimiter
3	Choose appropriate Delimiter
4	Click OK

Note: This works best where:

- Desired Output is Pivoted
- Number of elements will never change

It is a good idea to remove the auto-generated Changed Type step in case the number of columns changes on update.

Source

Week End	Values
Days	Monday
	Tuesday
	Wednesday
	Thursday
	Friday
	Saturday
	Sunday
2018-01-06	1
	2
	3
	4
	5
	6
	7
2018-01-13	12

Result

	Days	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
1	2018-01-06 12:00:0…	1	2	3	4	5	6	7
2	2018-01-13 12:00:0…	12	34	54	65	23	45	34
3	2018-01-20 12:00:0…	65	23	67	24	54	23	54

Splitting Data

Splitting Records into Rows

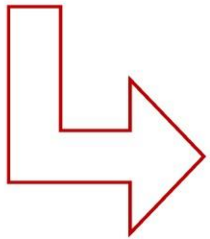
Step	Tasks to perform
1	Right click the Values column
2	Choose to Split Column →by Delimiter
3	Choose appropriate Delimiter
4	Click Advanced →Split to Rows →OK

Note: This works best where:

- Desired Output is Unpivoted
- Number of elements may change

Source

Date	Records
2018-01-31	Dog
	Cat
	Mouse
2018-02-28	Gerbil
	Horse
2018-03-31	Dog
	Cat
	Mouse
	Horse



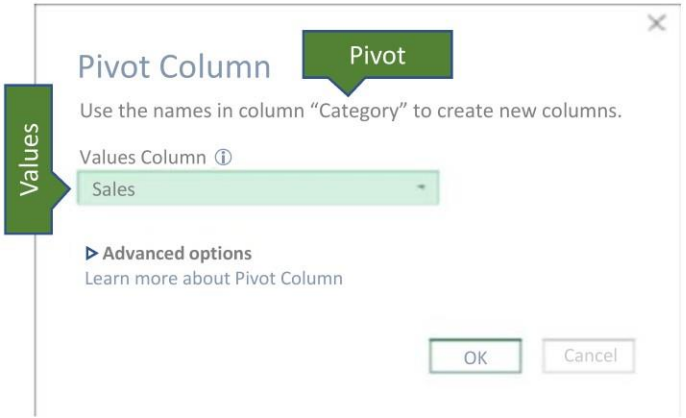
Result

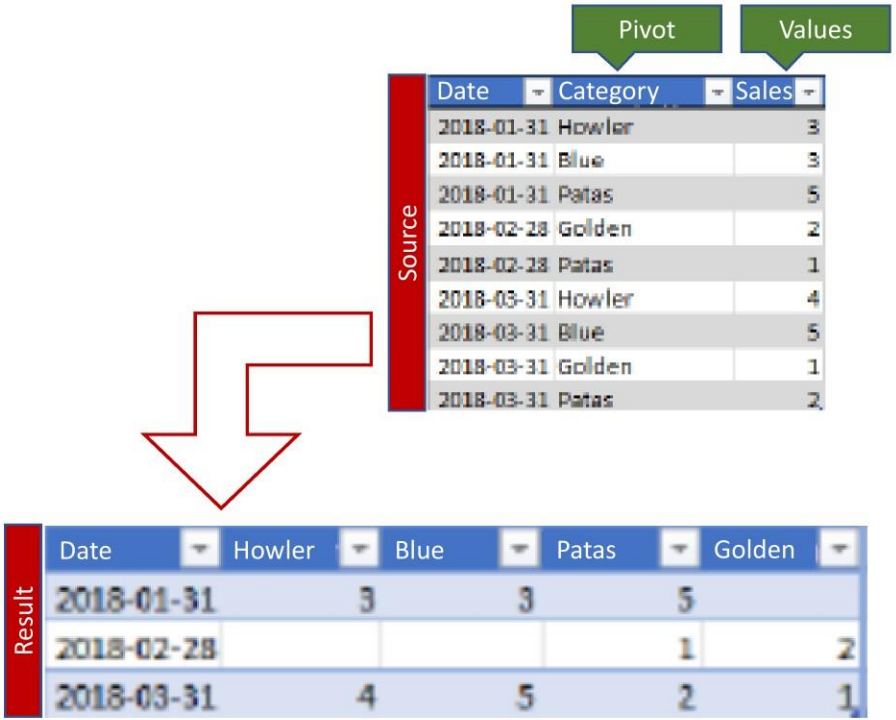
	Date	Records
1	2018-01-31	Dog
2	2018-01-31	Cat
3	2018-01-31	Mouse
4	2018-02-28	Gerbil
5	2018-02-28	Horse
6	2018-03-31	Dog
7	2018-03-31	Cat
8	2018-03-31	Mouse
9	2018-03-31	Horse

Reshaping Data

Pivoting Data

Step	Tasks to perform
1	Select the [Pivot] column
2	Go to Transform →Pivot Column
3	Select from the Values 'drop-down' column
4	Click OK





Reshaping Data

Transposing Data

Helpful Tips

Step	Tasks to perform
1	Prepare the Source table - Filter out Total and Subtotal rows - Delete any Total and Subtotal columns
2	Transpose the Data - Demote headers if you wish to keep them - Go to Transform →Transpose - Promote First Row to Headers - Set data types

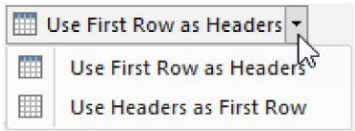
Source

Sales Category	Beer	Wine	Liquor	Cider	Cooler
2018-01-01	103	175	162	51	57
2018-01-02	243	223	207	60	69
2018-01-03	101	138	103	63	54
2018-01-04	137	57	179	51	69
2018-01-05	103	66	150	60	59

Result

Sales Category	2018-01-01	2018-01-02	2018-01-03	2018-01-04	2018-01-05
Beer	103	243	101	137	103
Wine	175	223	138	57	66
Liquor	162	207	103	179	150
Cider	51	60	63	51	60
Cooler	57	69	54	69	59

Note: To Demote Headers, go to …
Home →Use Headers as First Row
& To Promote to Headers, go to …
Home →Use First Row as Headers



Combining Data

7 ways to Join your Data

Inner Join

All Matched from both (no red, no yellow)

Full Outer Join

Everything from both Left & Right

Full Anti Join

All Unmatched from both (red + yellow)

Left Outer Join

Everything from Left (no yellow)

Left Anti Join

All Unmatched from Left (only red)

Transactions

Accoun	Dept	Date	Amount
64010	150	2015-12-15	8,975
64020	150	2015-12-15	13,708
64030	150	2015-12-15	32,555
64010	250	2015-12-15	22,752
64015	150	2015-12-15	34,147
64030	250	2015-12-15	19,733
64040	250	2015-12-15	33,438
64010	350	2015-12-15	45,876

Chart of Accounts

Account	Dept	Name
64010	150	Revenue - 18 Holes
64020	150	Revenue - 9 Holes
64030	150	Revenue - Twilight
64040	150	Revenue - Special
64010	250	Rentals - Power Cart
64020	250	Rentals - Pull Cart
64030	250	Rentals - Clubs
64040	250	Revenue - Golf Balls

Right Outer Join

Everything from Right (no red)

Right Anti Join

All Unmatched from Right (only yellow)

Combining Data

7 ways to Join your Data

1.  Inner Join
2.  Full Outer Join
3.  Left Outer Join
4.  Right Outer Join
5.  Left Anti Join
6.  Right Anti Join
7.  Full Anti Join

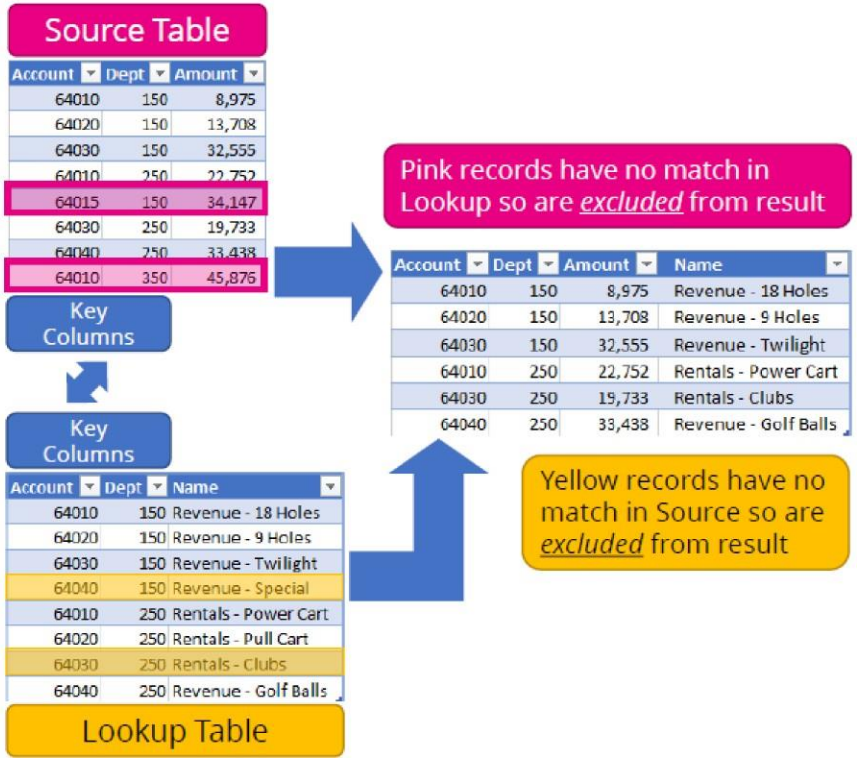
*Which of the **7** joins does each person fall under*

Combining Data

Inner Join

Step	Tasks to perform
1	For each table - Create a staging query (load as Connection Only)
2	Merge the Data - Reference the Source table - Go to Home → Merge → Lookup table - Select the [Key] column(s) on both tables
3	Expand the Data - Expand the new column - Uncheck the [Key] column(s) - Consider unchecking Preface option

Note: Use multiple columns for your [Key] column by holding down the CTRL key while selecting them.

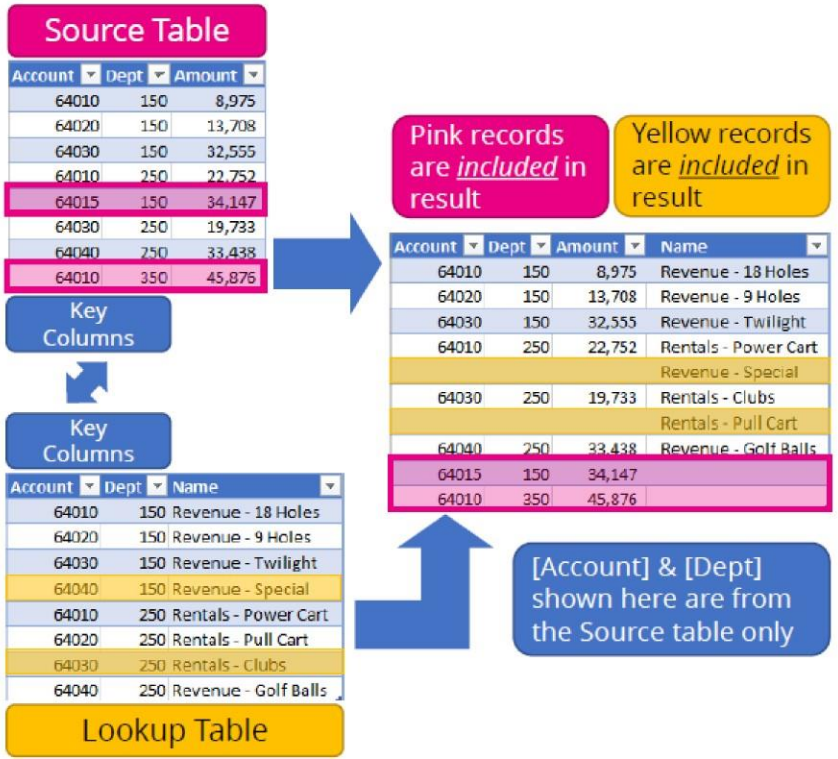


Combining Data

Full Outer Join

Step	Tasks to perform
1	For each table - Create a staging query (load as Connection Only)
2	Merge the Data - Reference the Source table - Go to Home → Merge → Lookup table - Select the [Key] column(s) on both tables - Use a **Full Outer Join**
3	Expand the Data - Expand the new column - Uncheck the [Key] column(s) - Consider unchecking Preface option

Note: Use multiple columns for your [Key] column by holding down the CTRL key while selecting them.

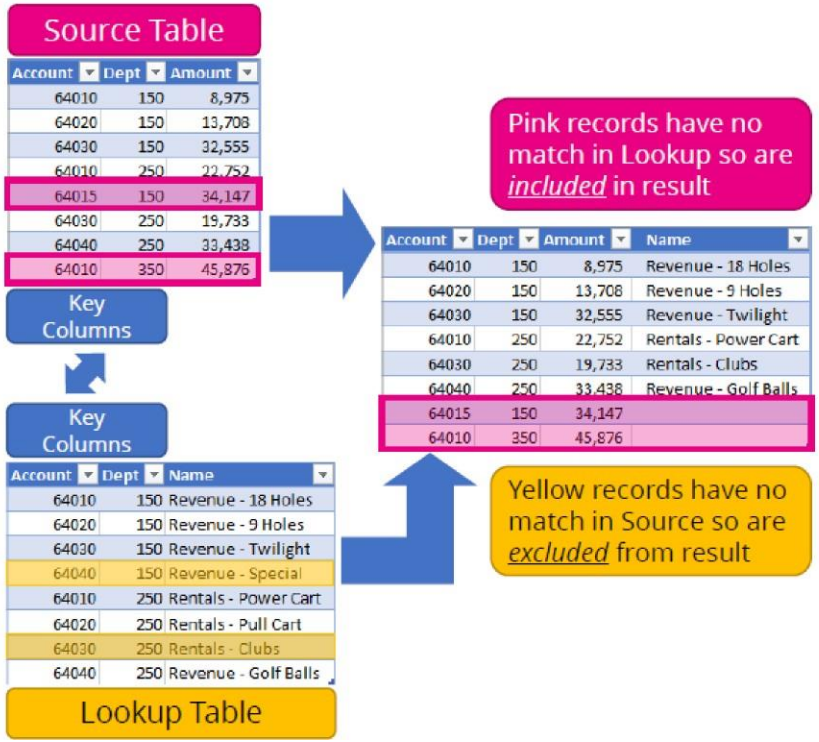


Combining Data

Left Outer Join

Step	Tasks to perform
1	For each table - Create a staging query (load as Connection Only)
2	Merge the Data - Reference the Source table - Go to Home → Merge → Lookup table - Select the [Key] column(s) on both tables - Use a **Left Outer Join**
3	Expand the Data - Expand the new column - Uncheck the [Key] column(s) - Consider unchecking Preface option

Note: Use multiple columns for your [Key] column by holding down the CTRL key while selecting them.

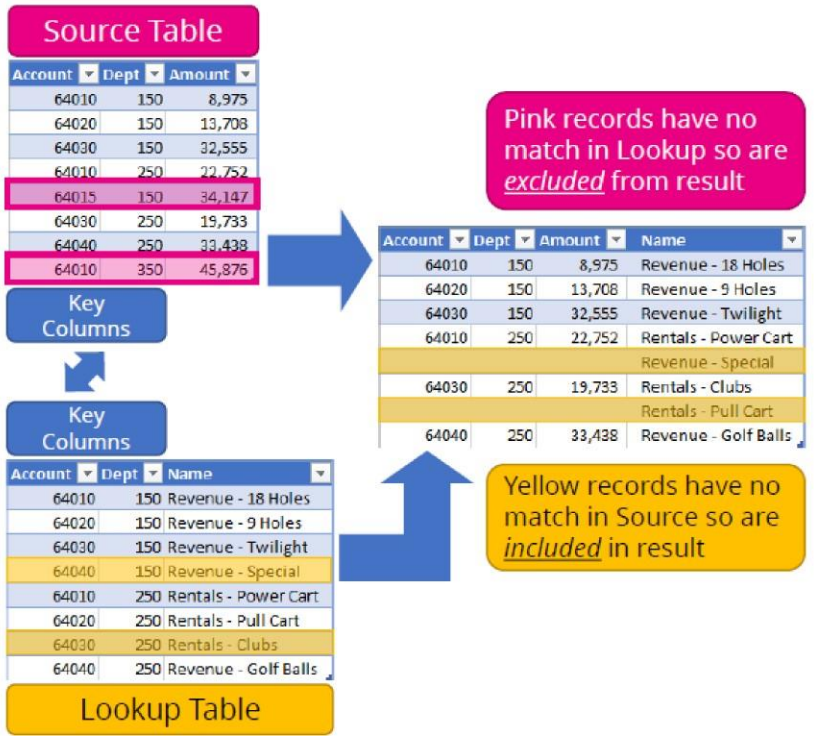


Combining Data

Right Outer Join

Step	Tasks to perform
1	For each table - Create a staging query (load as Connection Only)
2	Merge the Data - Reference the Source table - Go to Home → Merge → Lookup table - Select the [Key] column(s) on both tables - Use a **Right Outer Join**
3	Expand the Data - Expand the new column - Uncheck the [Key] column(s) - Consider unchecking Preface option

Note: Use multiple columns for your [Key] column by holding down the CTRL key while selecting them.

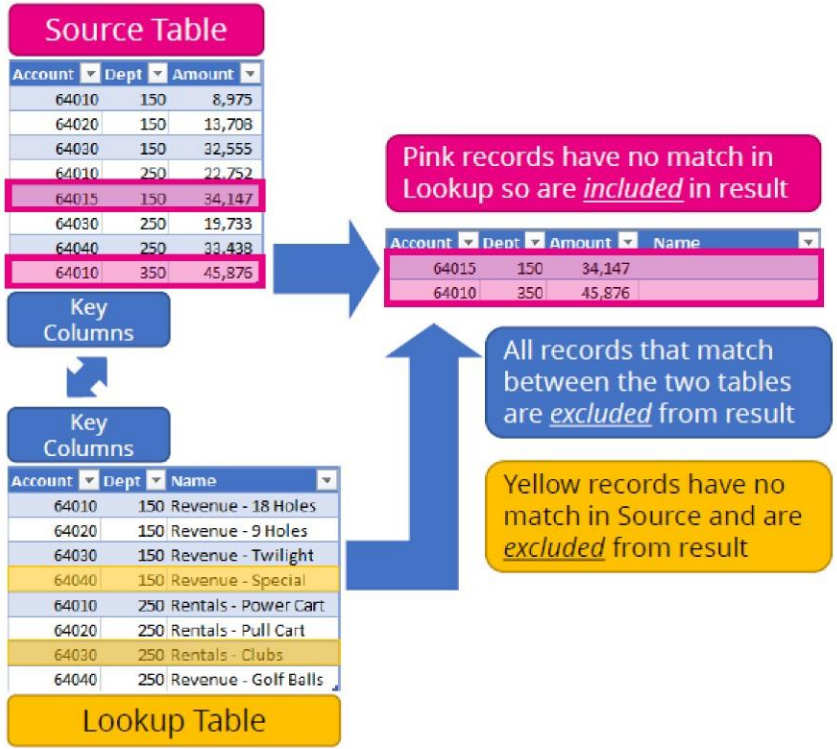


Combining Data

Left Anti Join

Step	Tasks to perform
1	For each table - Create a staging query (load as Connection Only)
2	Merge the Data - Reference the Source table - Go to Home → Merge → Lookup table - Select the [Key] column(s) on both tables - Use a **Left Anti Join**
3	Expand the Data - Remove the [Key] column(s) - Expand the new column - Consider unchecking Preface option

Note: Use multiple columns for your [Key] column by holding down the CTRL key while selecting them.

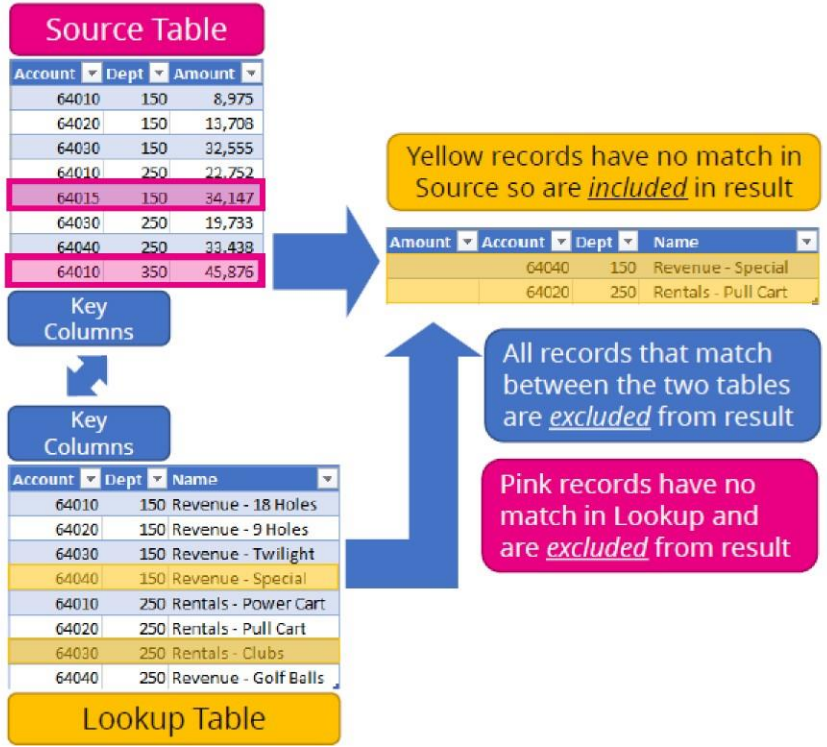


Combining Data

Right Anti Join

Step	Tasks to perform
1	For each table - Create a staging query (load as Connection Only)
2	Merge the Data - Reference the Source table - Go to Home → Merge → Lookup table - Select the [Key] column(s) on both tables - Use a **Right Anti Join**
3	Expand the Data - Remove the [Key] column(s) - Expand the new column - Consider unchecking Preface option

Note: Use multiple columns for your [Key] column by holding down the CTRL key while selecting them.

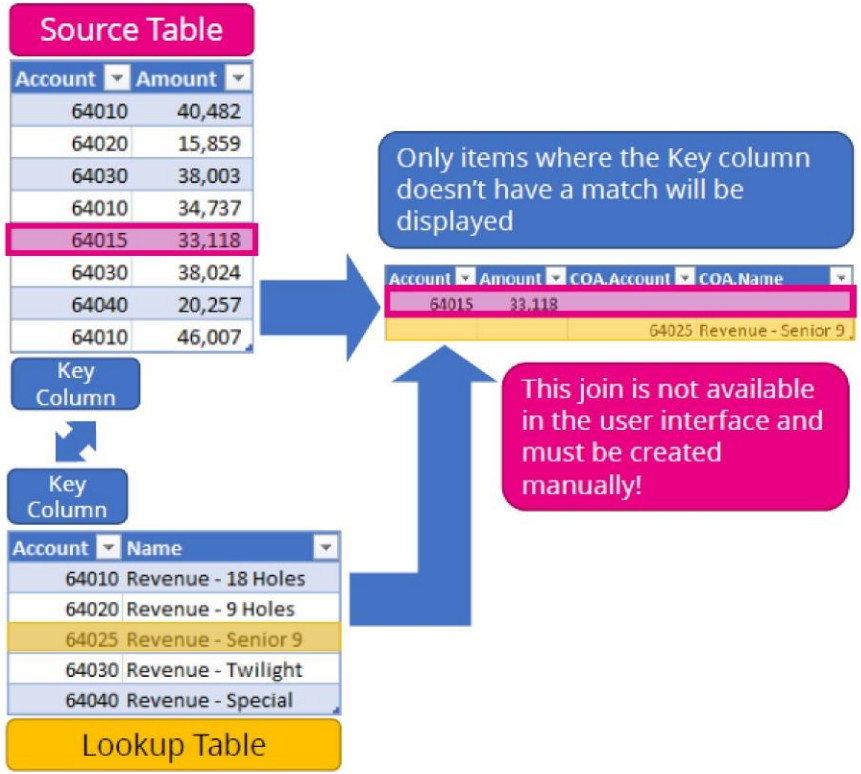


Combining Data

Full Anti Join

Step	Tasks to perform
1	Create Left Anti Join query
2	Create Right Anti Join query
3	Expand all fields from joined [Key] column(s)
4	Load as staging queries (Connection Only)
5	Reference Left Anti-Join query
6	Go to Home →Append →Right Anti-Join query
7	Load to your destination

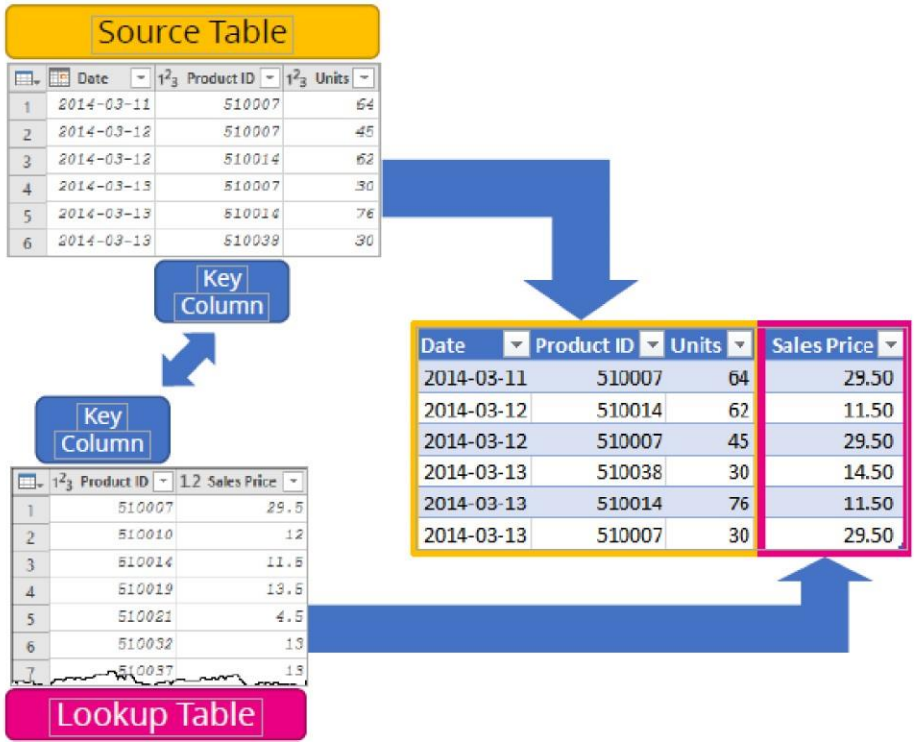
Note: Ensure names of columns are identical in both the Left Anti Join & Right Anti Join queries.



Combining Data

Exact Match

Step	Tasks to perform
1	Prepare Source table - Set data types - Load as Connection Only
2	Prepare lookup table - Set data types - Right click [Key] column →Remove Duplicates - Load as Connection Only
3	Perform the match - Reference the Source table - Go to Home →Merge →Lookup table - Select the [Key] column on both tables - Use a Left Outer Join - Expand the new column - Uncheck the [Key] column and prefix option



Appending Data

Basic Append

Step	Tasks to perform
1	Create a “Staging1” query - Create a query that connects to Data Source 1 - Make the required transformations - Load as Connection Only
2	Create a “Staging2” query - Create a query that connects to Data Source 2 - Make the required transformations - Load as Connection Only
3	Append the Data - Right click the “Staging1” query - Choose Reference - Go to Home Append - Select the “Staging2” query - Perform any additional required transformations - Load to your destination of choice