The Ultimate Power Query Cookbook for Power BI and Excel

Leveraging Power Query for collecting, combining and transforming your data

Dominick Raimato

www.bpbonline.com

First published: 2024

Published by BPB Online

WeWork

119 Marylebone Road

London NW1 5PU

UK | UAE | INDIA | SINGAPORE

ISBN 978-93-55517-395

www.bpbonline.com

Dedicated to

To the data community who strive to change humanity for the better.

Please continue to do the work you do for we would be lost without you.

About the Author

Self-identified as a data nerd, **Dominick Raimato** has had a huge passion for identifying insights and building solutions. You never know what kind of data projects he might be working on such as tracking temperature and humidity trends in his house to building a database of property assessments to make sure his house is in line with other properties to avoid overpaying taxes. Passionate about people, Dominick strives to create solutions that not just help achieve business objectives, but also make people's lives better. If he can help someone get home earlier to eat dinner with their family or eliminate time consuming tasks so they can focus on more important things, he is in his element.

Dominick currently lives in Bergen County, NJ where he can be found around the house working on projects or cooking in the kitchen. His specialty dishes include his homemade meat sauce in the winter and pulled pork bar-b-que on his smoker during the summer. He also ventures into New York City frequently to see his wife perform as a flutist with small ensembles, orchestras, and theaters.

About the Reviewer

Divit aspires to transform the landscape of Cloud Computing, AI, and Data Science by guiding a fresh cohort of professionals. As a senior member of IEEE, his objective is to equip individuals worldwide with advanced expertise in analytics, machine learning, and AI within the next five years, thus raising the global benchmark for technical proficiency. In alignment with the insights of influential thinkers, Divit contends that true accomplishment extends beyond personal success, emphasizing the importance of uplifting others and actively contributing to their pursuit of excellence. His overarching mission is to shape a future in the tech realm that is both skilled and inclusive.

Divit, an accomplished IT professional with a career spanning over two decades, is widely recognized as a prominent speaker and thought leader in the realm of emerging technologies. Hosting popular podcasts like "Tech Talk with Divit," "Live Labs with Divit," and "Cloud Bites with Divit," he provides a platform to showcase technological initiatives and leadership. With roles including Oracle TV's correspondent for Cloud World and a recognized expert, Divit presented on Oracle Database technology at Oracle Cloud World FY 2023.

Beyond his podcast endeavors, Divit is deeply passionate about knowledge-sharing, evident in his participation in international conference talks, technical blogs, Whitepapers, and the authorship of multiple books on emerging technologies. His expertise has gained visibility in leading newspapers and technology magazines worldwide. As a highly skilled technologist, Divit holds professional certifications from Microsoft, Oracle, AWS, Databricks, and has earned a post-graduate certificate from Harvard Business School Online.

Acknowledgement

I want to first and foremost express my gratitude and love to my family, friends, colleagues, and customers for your encouragement as I embarked upon this journey. Notably, thank you to my wife, Nicole, for your love and support every step of the way. I could not do any of this without you. I love you so much! Also, I promise to get back to doing the dishes more frequently.

I want to thank everyone at BPB Publications for their support and guidance through this entire process. Having never written a book before, I would be lost without your expertise during every step of the process. Thank you for the patience and collaborative support of all reviewers, experts, and editors through this entire journey.

I would like to thank my leaders at SHI, Matt Mallory and Brandon Sexton, for your investment in my success and supporting activites like this that allow me to grow.

I would like to thank Audrey Blair for helping me find my voice and take the leap as an active contributor to the Power BI community. Without your encouragement, this book would have just been an idea that would have sat idly in my mind.

I would like to thank Kristen Hubler and her aptly timed book *The Impossible Box* as it helped me get what was a dream out of storage and to go for it. Thank you for also helping me through the process and sharing the challenges of the book writing process.

I would like to thank Michael Heath for the challenges we face together. Having the opportunity to push each other to be the best version of ourselves is unique. Thank you for helping raise the bar each week. I look forward to reading your book one day too.

Finally, I would like to thank all of you reading this book! Your encouragement and support mean the world to me! I hope you enjoy reading it as much as I did writing it.

Preface

Data has infiltrated every aspect of our lives. The expectation of querying data and finding meaningful insights is now going beyond traditional business intelligence teams and becoming a required skill for analysts across the entire organization. While there are many tools out there that can be used to query and transform data, Power Query has proven to be a leader in this space.

This book is an all-inclusive guide to querying and transforming data with Power Query. Starting with accessing data from diverse sources, we walk through how data is transformed, cleansed, merged, appended, and manipulated to turn raw data into a meaningful data model to obtain insights.

Throughout this book, you will learn different ways transform your data to support efficient refreshes while preventing your model from experiencing bloat. We will explore best practices and use cases to help you understand how to best apply these concepts in your real-life scenarios.

This book is intended for anyone who is interested in data analytics or needs to manipulate data for their use. While the primary audience is for data analysts and business intelligence developers who build models for reports and insights, anyone who has a need to transform data will find value with the contents of this book. Whether you are a novice or expert, you will find value from the contents of this book.

With this book, you will obtain the knowledge and skills to query and transform your data queries used for data modelling within Microsoft Excel or Power BI. I hope you will not only learn the concepts in the context of this book, but find success in applying them in your work.

Chapter 1: Introduction to Power Query - Starting with the why, we will explore why Power Query is used to transform and manipulate data for analysis. We will then provide an overview of using Power Query both in Excel and Power BI. Next, we will review the interface and the value of using both the GUI interface and Advanced Editor for our transformations. Lastly, we will provide a brief guide on how to use this book to provide the most value with real life applications.

Chapter 2: Connect to File-Based Data Sources - We will walk through the process of connecting to individual files and multiple files in a single query both on a local device and through SharePoint. The goal at the end of this chapter is to enable all users of Power Query to be able to connect to at least one data source that they can use to complete their

training without having to leverage additional resources that might cost additional money. We will also discuss the use of the Power BI Enterprise or Personal Gateway as appropriate.

Chapter 3: Connect to Web-Based Data Sources - Starting with simple web connections like SharePoint lists and tables on web pages, we will pull in data that can be manipulated within Power Query. From there, we will go through the steps to pull in data using web APIs using the web connector and OData connector. We will then use the Web Connector to pull data from a SharePoint list to filter the data coming in and optimize the query. Lastly, using all the concepts covered up to this point we will connect to data in Dynamics 365.

Chapter 4: Connect to Database Sources - We will start with the common SQL-based connections (MSSQL, MySQL, PostgreSQL, etc.) to establish a solid foundation for connecting to databases. From there, we will jump into more advanced and nuanced connectors. We will discuss the differences, challenges, and some considerations with different connectors like Snowflake, Google Big Query, ODBC, Dremio, Redshift, and Analysis Services. Lastly, we will connect to Microsoft's Dataverse and Dataflows and bring them into Power Query.

Chapter 5: Connect to Third-Party Data Sources - We will go through the basics on how to connect to common SaaS platforms for deeper analysis. We will discuss some of the challenges associated with connecting to them as well as some potential additional costs that might be associated with them. While these connectors tend to be a little easier to manage, the goal is to provide a little more context with them including some potential gotchas that might come up during the connection process.

Chapter 6: Managing Data Types - We will dive into the use of data types in Power Query. While this topic may seem straight forward, there are a few implications that need to be considered when selecting data types as not are created equal. How and when you use data types can have implications on the data model and can contribute to bloat. In addition, downstream effects can be had upon the model when queries are merged and even relationships in the model. We will end this chapter talking about best practices when using data types to provide the most efficiencies with your model.

Chapter 7: Transforming Columns - We will review the different ways data can be transformed for use. One of the biggest challenges when working with data is making sure that the data is in a format that we can use. Sometimes these transformations are basic like updating the column headers to something that makes sense while other times we are looking to extract substrings from a long line of text. Regardless of your specific needs, we are able to apply this logic easily in Power Query. The goal is to apply your transformations within Power Query so they are automated as you ingest new data into your model when refreshed.

Chapter 8: Cleansing Columns - We will take transformation another step further and go through the process of cleansing your data. While we would like to think our data is clean and ready to use, we often find issues with our data sets. It could be as simple as a few extra rows at the top of an Excel spreadsheet before the actual data is made available. Sometimes we have errors in cells or even blanks that need to be handled. We also get extra columns that might not be necessary to do what we need to do. As a result, we can easily cleanse the data with Power Query to get everything into shape. Just like the chapter before, the goal is complete automation. This will allow us to never perform these tasks again when we bring new data into the data model.

Chapter 9: Creating New Columns - We will go through the process of creating new columns for our query. We will start by briefly reviewing what transformations we saw in Chapter 7 can be turned into new columns while preserving old columns. We will then dive into splitting columns by delimiters. We will then go through the process to manually create new columns starting with the conditional column tool. From there, we will show you how to create one with your own custom formula. Then we will show you how to save time and energy by creating from example columns if you are not comfortable with writing your own formulas. We will then introduce the concept of the Index and Function columns but reserve the details for later chapters. Lastly, we will go over how to streamline your queries by updating the data type for a custom column using the Advanced Editor.

Chapter 10: Combining and Manipulating Queries - We will talk about transforming the table as a whole. Starting with a quick review of aggregations, we will quickly apply them to the Group By function to help streamline the query and reduce the number of rows. We will also learn how perform aggregations on specific columns using the statistics function that can be used in custom columns. From there, we will go through the process of manipulating tables for better use. Frequently, the data we receive is nicely formatted for use in a tool like excel but does not work well for modelling. We will use the Transpose and Pivot tools to get the data into a usable shape for our analysis.

Chapter 11: Using Python, R, and AI - As a result of your analysis of a data model, you might find you have enough data to create a machine learning model or desire to use R or Python for deeper analysis. Once you have a working model, you might want to integrate them into Power Query to perform predictive analytics. In this chapter, we will discuss the integration of AI, R, and Python in a data model with Power Query. The focus of this chapter will not be on the creation of AI models as that is an entire book unto itself but rather how to apply them in Power Query. This is always a top ask in any Power Query training course and it is important to understand how these tools are leveraged as many think they can apply AI with a few clicks.

Chapter 12: Indexing - We will explore how we can leverage the index column to help establish our star schema in our data model. We will start with a quick review of the Star Schema and why it is the model of choice for Power Pivot in Excel and Power BI. We will then go over what keys are in a data model and what kinds we might use. We will dive into creating a primary key on a dimension table using the index column. From there, we will go over how create a foreign key on a fact table using some advanced formulas within a custom column. Lastly, we will discuss the value of creating alternate keys in Power Query to help optimize our models and simplify their use.

Chapter 13: Parameters - We will review why parameters are a valuable tool when it comes to transforming your queries. To me, considered to be one of the most transformational tools in Power Query, parameters are often ignored and rarely are used to their full potential. We will dive into how we can use parameters to simplify our connections and manage multiple queries at once. We will also dive into how they are a critical component in setting up Incremental Refresh for Power BI. Lastly, we will go over how parameters are helpful for template files and with the Power BI service.

Chapter 14: Functions - We will work to demystify functions in Power Query. Functions allow us to apply repeatable transformations or load data quickly with only a few clicks. What is beneficial about them is that we can create the functions and load them into a template for use with minimal overhead. This allows us to quickly deploy them over and over again without having to search for the code we used before. We will go through the process of taking a query and converting it into a function so it can be deployed again. We will also go through the creation of two common use cases I use all the time for functions – converting time zones from UTC to local time and creating a date table.

Chapter 15: Advanced Web Connections - We will go through the process of creating advanced web connections with Power Query. Using endpoint parameters, we will filter our connections to streamline refresh times. We will also investigate the use of dynamic content such as date ranges and dynamic connection strings to ensure sustainability with our queries. We will also go through the process of creating a function to acquire an OAuth2 Token and leverage it with other queries. Lastly, we will discuss some of the common challenges and limitations you might run into when it comes to using web connectors in Power Query.

Chapter 16: Manipulating Supporting Queries - We will go through the process of manipulating supporting queries for connectors that automatically append files together. Applying specifically to the folder and blob connectors, it is important to understand how changes need to be made in these supporting queries to avoid any rework. This allows you to continue automating your queries without requiring any additional manipulation outside of Power Query. By automating this process, this will simplify future additions to your data model.

Code Bundle and Coloured Images

Please follow the link to download the
Code Bundle and the ***Coloured Images*** of the book:

https://rebrand.ly/2anl7bu

The code bundle for the book is also hosted on GitHub at

https://github.com/bpbpublications/The-Ultimate-Power-Query-Cookbook-for-Power-BI-and-Excel.

In case there's an update to the code, it will be updated on the existing GitHub repository.

We have code bundles from our rich catalogue of books and videos available at **https://github.com/bpbpublications**. Check them out!

Errata

We take immense pride in our work at BPB Publications and follow best practices to ensure the accuracy of our content to provide with an indulging reading experience to our subscribers. Our readers are our mirrors, and we use their inputs to reflect and improve upon human errors, if any, that may have occurred during the publishing processes involved. To let us maintain the quality and help us reach out to any readers who might be having difficulties due to any unforeseen errors, please write to us at :

errata@bpbonline.com

Your support, suggestions and feedbacks are highly appreciated by the BPB Publications' Family.

Piracy

If you come across any illegal copies of our works in any form on the internet, we would be grateful if you would provide us with the location address or website name. Please contact us at **business@bpbonline.com** with a link to the material.

If you are interested in becoming an author

If there is a topic that you have expertise in, and you are interested in either writing or contributing to a book, please visit **www.bpbonline.com**. We have worked with thousands of developers and tech professionals, just like you, to help them share their insights with the global tech community. You can make a general application, apply for a specific hot topic that we are recruiting an author for, or submit your own idea.

Reviews

Please leave a review. Once you have read and used this book, why not leave a review on the site that you purchased it from? Potential readers can then see and use your unbiased opinion to make purchase decisions. We at BPB can understand what you think about our products, and our authors can see your feedback on their book. Thank you!

For more information about BPB, please visit **www.bpbonline.com**.

Join our book's Discord space

Join the book's Discord Workspace for Latest updates, Offers, Tech happenings around the world, New Release and Sessions with the Authors:

https://discord.bpbonline.com

Table of Contents

Chapter 1
Introduction to Power Query

Introduction

This chapter will provide an overview of Power Query. We will cover the basics of why Power Query is a useful tool and what value it provides for data analysis and reporting. Next, you will learn how to leverage Power Query inside of Microsoft Excel or how to prepare your device to use it inside of Microsoft Power BI. We will then provide an overview of the graphical interface for Power Query and cover the basics of leveraging the M Query language inside the advanced editor. Lastly, we will review how best to leverage this book for success with your projects so you can extract the most value when using Power Query.

Structure

In this chapter, we will cover the following topics:

- Why use Power Query?
- What is Power Query?
- Leveraging Power Query with Excel
- Leveraging Power Query with Power BI Desktop
- Overview of the Power Query graphical interface

- Overview of the advanced editor
- Maximizing Power Query

Objectives

By the end of this chapter, you will understand the value and purpose of Power Query for your analysis and reporting projects. You will be able to set up your device to follow along with the topics found within this book and complete the steps outlined. As you expand your knowledge with Power Query, you will also understand the best way to extract the most value from this book.

Why use Power Query?

In 2015 there was a gentleman, Andrew, who worked in finance and performed reporting for the quality department of his company. For years, he had put together reports for department leaders so they could review their monthly budgets and see if they had met their forecast plan. These reports aimed for cost center leaders to see how they performed against their plan. They received these reports a few days after the financial period closed in a Microsoft Excel file so they could review the details of their actuals and adjust their forecasted spend for the remainder of the year. It seemed like a straightforward task that could be completed quickly.

We could not have been more wrong about how difficult this task was to complete. Andrew shared his story about spending three days assembling ten reports so these department leaders could review them. In surveying the cost center leaders, most of them only spent a total of five minutes reviewing these reports. In addition, a few leaders admitted that they never actually opened them. Andrew spent a lot of time and effort on something that was barely touched by his target audience.

After spending time with Andrew, we understood the core of his challenge. He exported data from their accounting system and needed to transpose it into a format that was easy to consume. But every time he exported the data, he had to complete multiple transformations to force the data to fit into the desired template. He performed the same transformations for each cost center he supported, so he repeatedly spent hours putting together each report for a single department.

Andrew's challenge is not unique. If you are reading this chapter, you likely have a challenge that is quite similar. You have reports and analyses that need to be completed, but you spend hours repeating the same steps. You have wasted hours preparing data for use or have felt the frustration of doing the same thing repeatedly and wondered if there was a better way.

Well, if you feel like Andrew, then there is hope! There is a better way with Power Query! Power Query allows us to apply transformations on our data that can be repeated. By

spending a little extra time upfront, we can automate our transformations as we add more data to our models. This process can be daunting at first, but given some time and practice, you will find it easy to do with your data.

One of the benefits Andrew found besides reclaiming three days of his work was his ability to become more proactive with his reports. With three days reclaimed, he could spend time performing meaningful analysis and providing insights to department leaders. He had always wanted to do this but struggled to complete it because he did not have the bandwidth to do so. And this was just the beginning for Andrew, as once he could automate the transformations, he could deliver the reports more frequently. This allowed Andrew to enable cost center leaders to become more proactive with their budgets. With reports coming in on a weekly basis, these leaders could determine if they were going to overspend their forecast and correct course before it became a larger problem. Andrew could never deliver their reports weekly because that would have become his full-time job. However, by simplifying his workload with Power Query, he became a better partner with other leaders in the business and transformed how financial reporting is completed.

While Andrew's scenario might not apply specifically to you, there are multiple use cases where Power Query can prove to be useful. Instead of exporting data from a system, you might connect directly to the underlying database to acquire the data instead of performing an export. Or maybe your initial data source only paints part of the picture, and you need to combine it with another source to provide the full context. As we explore together through this book, we will identify these types of scenarios and navigate them with the assistance of Power Query.

Whether you are a seasoned data professional or a beginner, you probably will encounter someone who will say something along these lines about Power Query: "Why would you perform transformations in Power Query? You need to perform your transformations at the data source. No data professional would ever perform their transformations with Power Query!" As a best practice, we want to provide as much transformation upstream as possible. In a perfect world, every person would have access to a data warehouse and a data engineer to provide transformed data upstream from Power Query.

As wonderful as this sounds, that is just not the reality of the world we live in. Many of you reading this book might not have access to perform transformations on your data at the source. You might have received data in the form of a spreadsheet or a comma-separated value file. Maybe you are connecting to an **Application Programming Interface** (**API**) or using a connection to a third-party software service. In that case, all your transformation will take place inside of Power Query.

In the end, Power Query is designed to simplify the transformation of your data so you can focus on the analysis and insight of your data. That is far more valuable than spending time repeating steps to add new data to your model.

What is Power Query?

Now that we understand why Power Query is a valuable tool, let us talk about what it is. Power Query is an extraction, transformation, and loading tool. You sometimes hear it referred to as an ETL tool. Developed by Microsoft, Power Query is designed to allow users to extract, transform, and load data from disparate data sources and combine them into a single data model. Since many people have Microsoft Office installed on their computers, they can start leveraging Power Query today by opening Microsoft Excel. Others have elected to use a more robust visualization tool, such as Microsoft Power BI, to display their data but will leverage Power Query to get data into their model.

The first step with any project using Power Query will be connecting to and extracting data from various sources. You can connect with a diverse set of data sources ranging from files, databases, and web-based APIs. Some third-party software providers have created their own connectors inside of Power BI, such as LinkedIn, SurveyMonkey, and Zendesk. The best part of Power Query is that it allows you to connect to various data sources all at the same time. This enables you to connect data from different sources into a single model. For instance, you could connect your online sales data inside Dynamics with your customer service data found inside Zendesk. While the systems are different, we can bring the data together to find better insights.

Once we have extracted our data, we are able to perform transformations on it. Through a series of steps, we can go through and make changes to the data we have extracted and make the transformations repeatable. When using Power Query, we can refresh the data whenever we want, so we want to re-apply the same transformations each time we add data to our model. This is where most of the effort is put in to optimize our time in the future, as everything we do now will cascade down for future refreshes.

Once we have completed our transformations, it is time to load our data. When using Power BI, we will be using a data model. However, we have some choices when it comes to Excel. We can load the data we transformed into different sheets inside the file. However, we can load our data into the data model found inside PowerPivot. Instead of storing the data inside the cells, we can leverage it from the model instead.

This is where the most exciting part comes in regarding leveraging Power Query. We can manage large volumes of data with no issues. When performing analysis with Excel, we are limited to one million rows of data. However, if your machine is not properly configured, your performance might be severely degraded, resulting in a lower threshold for your row limit. However, Power Query paired with a data model can handle billions of rows with no issues assuming the resources are available to handle that volume.

As we go through this chapter, you will understand how Power Query serves as an ETL tool and simplifies how you prepare your data for use.

Leveraging Power Query with Excel

Microsoft Office serves as the standard when it comes to productivity software for most businesses. As a result, you likely have access to Power Query already with Microsoft Excel. To leverage Power Query, go to the data tab and expand the get data menu. From there, you can select a data source to get you started or just launch the Power Query editor directly from the menu. Refer to the following figure:

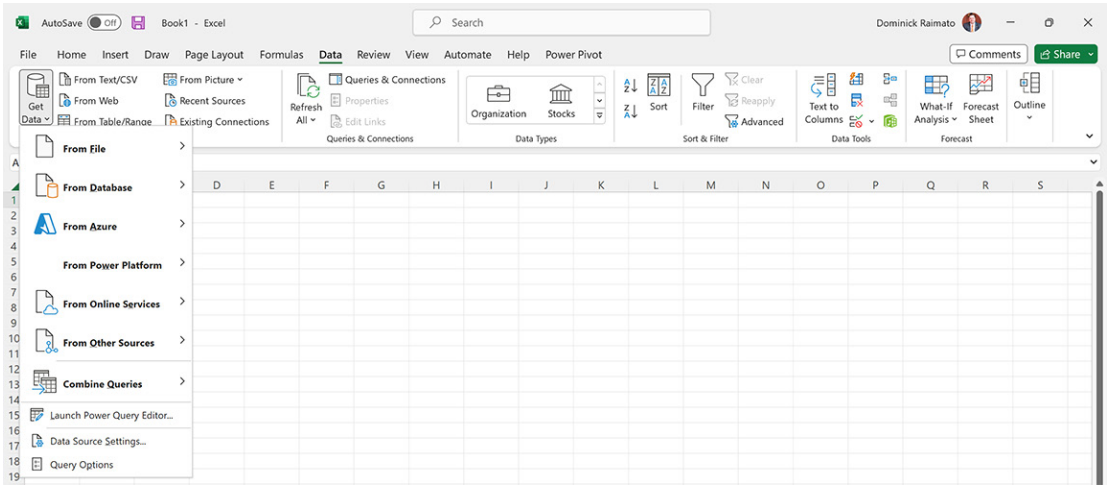

Figure 1.1: *Opening Power Query in Excel*

If you have a current version of Microsoft Excel already installed on your machine, then you are most likely ready to start using Power Query. Power Query has been available since 2010 but requires an add-in to be installed. In the 2016 edition of Excel, it was integrated with the tool and has been available with every subsequent version. Hopefully, you will be using a newer version of Excel as both the 2010 and 2013 versions are no longer supported, but keep in mind if they are still in use.

Since Power Query is best known for being leveraged with Power BI Desktop, users often ask if there is any advantage to using Microsoft Excel. The short answer is yes, as Excel serves a different purpose than Power BI. For example, a common use case for Power Query in Excel is to cleanse and transform two separate lists into a single table so it can be referenced by a script in PowerShell. You can perform the transformation with Power Query and load it into an Excel spreadsheet, so it is ready to be imported. Always remember that it is important to use the right tool for the right job.