ARTICLE

on

# DATA PREPARATION WITH POWER QUERY

Submitted To

MENTORNESS



IT Services and IT Consulting

Ahmedabad, Gujarat

In Partial Fulfilment –

Of Data Analyst Internship

Submitted by-

VIRAMGAMA DARSHAN KAMLESHBHAI

Pursuing MBA in Business Intelligence

BKBI, Gujarat University

Ahmedabad, Gujarat

# DATA PREPARATION WITH POWER QUERY

## What is a Power Query?

Power Query is a powerful data transformation and data prep engine. It's an essential part to do data acquisition regardless of the use case of your data. It allows users to connect to many different data sources, whether it be a database, an excel file, a web service, or cloud storage it can connect you to your data.



Here are the primary pieces that encompass Power Query.

### 1. Data Connectivity:

Power Query allows you to seamlessly connect to a massive amount of data sources, no matter if it's an extremely large data set, or a data source that isn't shaped how you need in order to be used Power Query can connect to your data.

### 2. Transformations and Shaping:

Power Query editor allows users to apply many different transformations to their data. This GUI is built in a way that it allows users to interactively build their queries, apply filters to their data, merge tables, and shape their data in other important ways.

The operations you define in a Power Query are repeatable, so that the data can be refreshed as long as needed.

### 3. The ETL (Extract, Transform, Load) process:

Power Query performs the whole ETL process: it extracts data from the source, transforms it to meet your requirements, and loads the data into your destination (such as Power BI).

# Why Is Power Query Essential for Power BI?

**Clean Data for Visualizations:**

For visualizations, Power Query helps you avoid misleading errors in data.

Before any visualization appears: With Power Query your data are cleaned up and transformed. You'll no longer have to worry about graphs that 'lie' or misleading tables in Power BI. There's too much 'tell' in 'showing'.

Simply put, Power Query is at the intersection of raw data and insightful visualizations. It flexes the muscles of those using it to more efficiently manage their data and thereby makes Power BI an even better tool for understanding and analyzing data.

## How Power Query Helps :
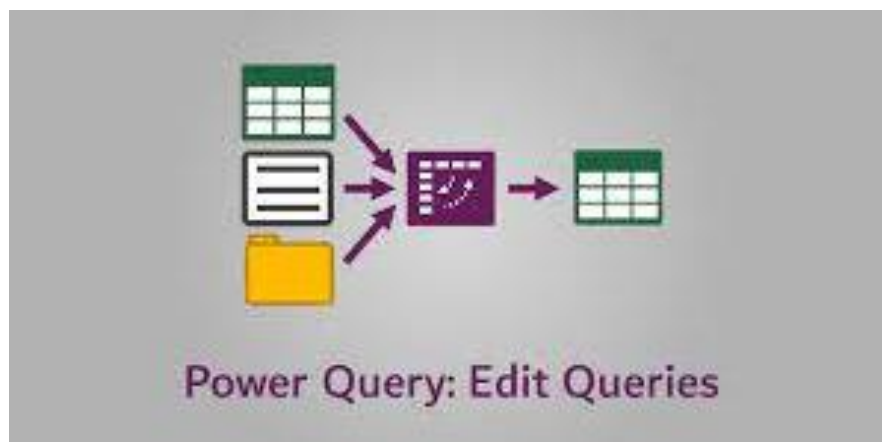


1. **Simplifying Data Acquisition:**

   Business users often spend a lot of time on data preparation. Power Query combines data of varying data formats to get consistent connectivity experiences from data sources.

   Whatever databases or irregular files you're handling, Power Query provides you a unified experience.

2. **Interactive and Intuitive Shaping:**

   Power Query 's interface is interactive, and allows users to build queries over just about any data source. You can easily modify the data structure, filter rows and create calculated columns.

   Changes you make during the reshaping process will always be reflebted in the future, keeping the same

3. **Handling Data Volume and Variety:**

Power Query lets you work against a subset of the entire dataset. You can define the data you need and perform the necessary transformations on it. At the same time, Power Query won't overload your system.

Power Query adapts, whether your data amounts to only one email or 10 trillion bytes, and whether it arrives from Hive at 1 msg/day or AMPLab at 100,000 messages/millisecond.

4. **Scheduled Refresh and Programmability:**

Power Query queries can be manually refreshed, or automatically scheduled (e.g., in Power BI).

To refresh your data now, you can use the Excel object model or other tools programmatically.

## COMBINING QUERIES

### Appending Queries in Power Query

**Appending queries** is a prominent feature in **Power BI** that allows you to create a joint table by combining similar data tables. Imagine it as stacking one table on top of another.

Here's how it works:

#### 1. What does Appending?

Into one query combine the results for two or more queries, each represented by a table.

Adding one table to another top-down is akin to appending in full result set of queries on a single table.

#### 2. How does Appending Work?

When you have multiple data sources with similar structures (e.g., daily sales data for different months).

With the inclusion of these all of these queries into one concatenated dataset, you have an unified data one.
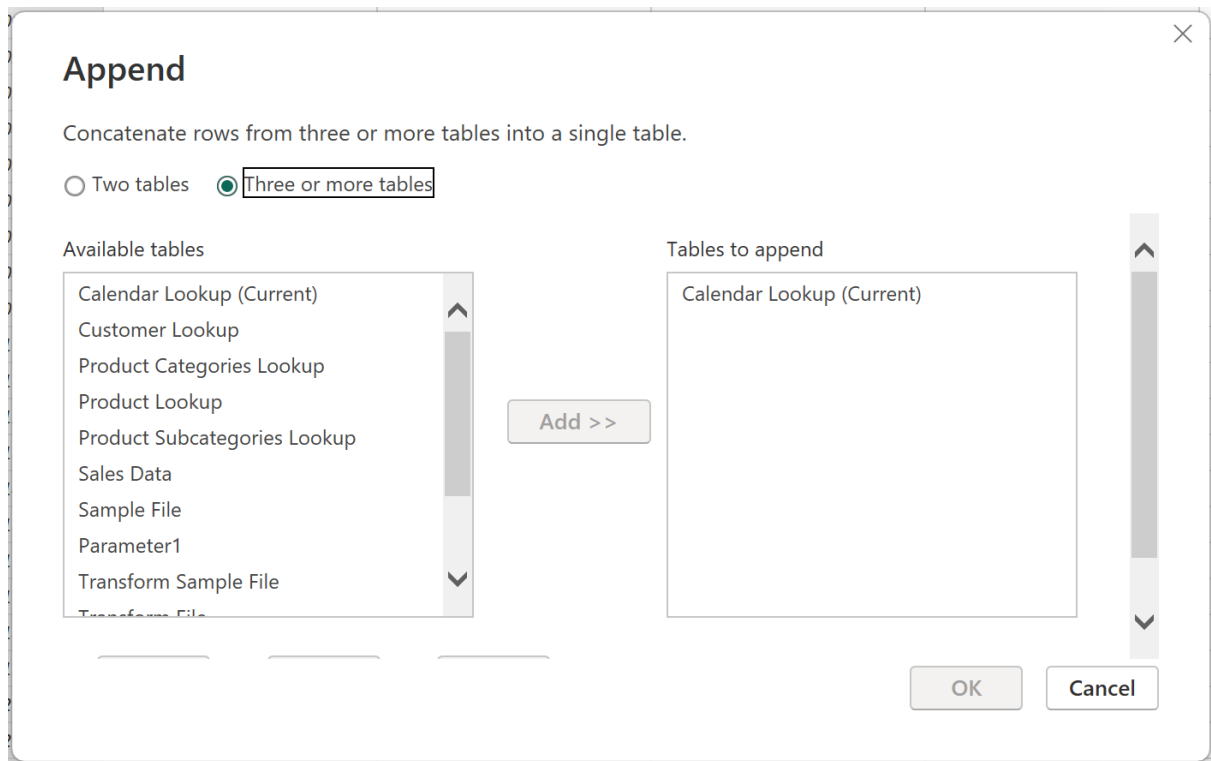
**3.How to Append Queries:**

Suppose you have two datasets Sales Data 2020 & Sales Data 2021.

- Append Queries from the Home tab.
- Choose Queries to Append.
- If you want to append only 2 table than choose another table.
- But if you want to append 3 or more than 3 than choose '3 and more table' option next to it and than bring all those table which you want to append to the right.(make sure all selected table has the similar column)
- Than press "Ok"
- You will be getting a new table.

## Merging Queries in Power Query

Another important skill is merging queries into some kind of a useful form. Continuing with this introduction to Merging:

1. **What is Merging?**

   By merging data or tables along their shared column(s), merging operations bind various datasets or tables together horizontally.

   Imagine plugging a series of different tables into each other to obtain a third table.

2. **Types of Merge:**

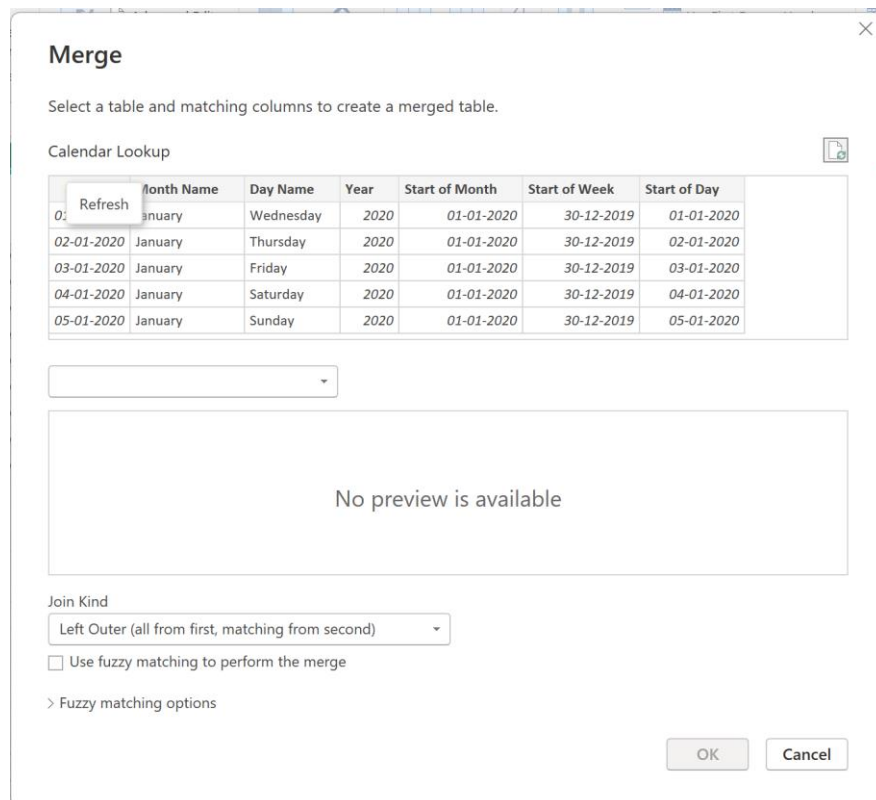Inner Merge: Rows which match in both tables are retained.

Left Outer Merge: All rows from the first table are preserved and only the rows from the second table which match are taken.

Right Outer Merge: Takes all of the rows from the second table and only the rows from the first table which match.

Full Outer Merge: Every row in both tables remains. Some missing values are filled in where a value is needed.

Left Anti Join: Left Anti Joins in Power Query only include rows in the left table where there are no matching rows in the right table.

Right Anti Join: Alternatively, in Power Query a right anti join returns only rows from the right table that do not match any rows in the left table.



### 3. How to Merge Queries:

Suppose we have Sales Data and Product Data.

Steps:

- In the Power Query Editor, select the Sales Data table as your base query.
- Click on the Home tab then click Merge.
- Specify (e.g.,Product Key) is the shared column between the two tables.

- Choose the second query column (e.g., Product Data) and corresponding to choose the first.
- Click OK to merge the queries.

Remember:

Merging queries will change the existing table, and merging queries as a new makes such a table out of the combined two or more. (Likewise, Append Queries)

You can choose several columns to join together into a dataset.

To sum up, no matter whether it is appending or merging that we are doing here, both are necessary operations for preparing your data within Power BI. No matter if you are stacking like tables on top of one another or taking data from different columns for orderly assembly; mastering these methods is sure to help make lighter your journey upward through data analysis!

## TRANSFORM & ADD COLUMN

In Power Query, the "Transform" and "Add Column" features are both essential to effectively manipulating data. Even though both share many features, there is considerable difference between them.

Two features are for directly modifying data within your dataset, namely: renaming columns, changing data types, filtering rows, sorting data, and applications of other transformations.

Their primary feature is really about managing new data or existing columns in the dataset.

But "Add Column" extends the dataset by creating new columns through calculations or transformations based on existing data.

In short, "Transform" changes existing data, while "Add Column" takes the dataset and adds new information that is extracted from existing pieces. These two functions provide essential tools for preparing and analyzing data from power query. You will likewise have more flexibility with your dataset's structure as well as its content, providing you with the means to adapt it at will.

So lets discuss some important feature of Transform and Add Column:
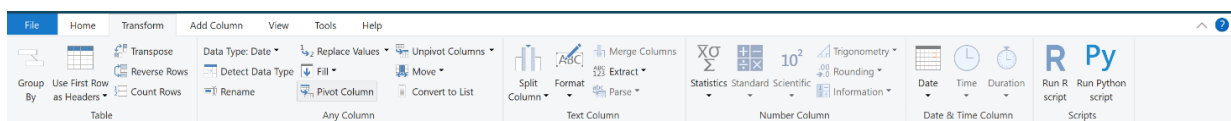
**Replace Value:**

With this operation, you can swap out specific data in your dataset with others. To illustrate: Null values or incorrect data entries, or any other values that you want to get rid of shouldn't be hard to deal with using Power Query's "Replace Values" function.

**Pivot and Unpivot Columns:**

Pivoting changes unique values in a given column to separate columns. Typically this is done to summarize data. Unpivoting, on the other hand, goes in the opposite direction--it takes many columns and turns them into two: one for attribute names and another with their respective values. Actually, you may want to respin your data for analysis or reporting purposes.

**Split Columns:**

Doing this operation will let you take a single column and chop it up into many columns based on a delimiter(such as a comma or space). It's handy when you have data stored in a single column that needs to be separated into distinct categories.

**Merge Columns:**

It can also work in the opposite direction, taking data from two or more columns and putting it all into one. This can be handy in cases where you want to sort things out or try joining strings from different columns.

**Standard and Scientific Numerical Operation:**

Power Query offers up a whole selection of basic plus scientific calculations like addition, subtraction, multiplication and division. What's more, it even supports scientific operations too, including exponentiation, logarithms, trigonometric functions as well as statistical functions such as mean, median, mode, standard deviation, etc.

**Statistical Tools:**

Power Query now has a variety of statistical functions. They range from measures of central tendency (mean, median, mode), to measures of dispersion (standard deviation, variance), measures of association (correlation, covariance) and so on. These tools tell you where data stands in terms of distribution and relationship.

**Date and Time Columns:**

Power Query offers various functions for modifying date and time columns. Moreover, you can extract components like year, month, day, hour, minute, and second from datetime values. Moreover, you can compute operations such as the difference between two dates by using an expression; add or subtract days, months, years; and reformat dates in various styles.