# Combining Data

## Why should I combine queries?

* Data will be in multiple sources and there is insight to gain it together

## Quote

It is perfectly reasonable to expect that your tables will have a consistent format before you combine them together.

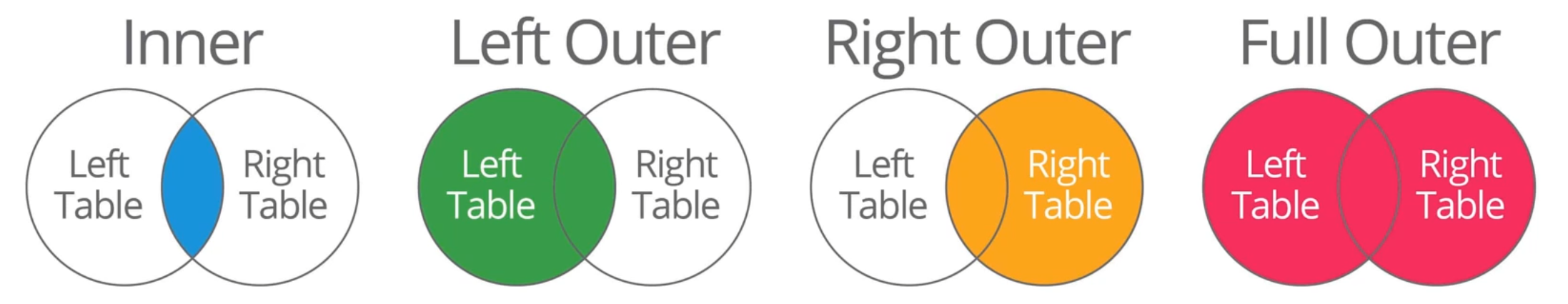
However, although this may be a reasonable requirement in ideal scenarios, it is selfdom feasible in real life.

You may need to work with tables that have different formats or, even worse, that may gradually become different.

## Two Ways to Combine

1. Combine DIFFERENT tables via MERGE
   1. Extend insight
   2. Example a student had different pieces of information about them. Maybe it had their grade history, club participation, maybe it had their study record. You might want to merge all of that information from those three different sources into one view of that student.
2. Combine LIKE tables via APPEND
   1. More of the same
   2. Adding more of the same information into a singular view. If we wanted to combine the grade for all of the years they have been at BCIT. We would use append to combine them into one view of that student.

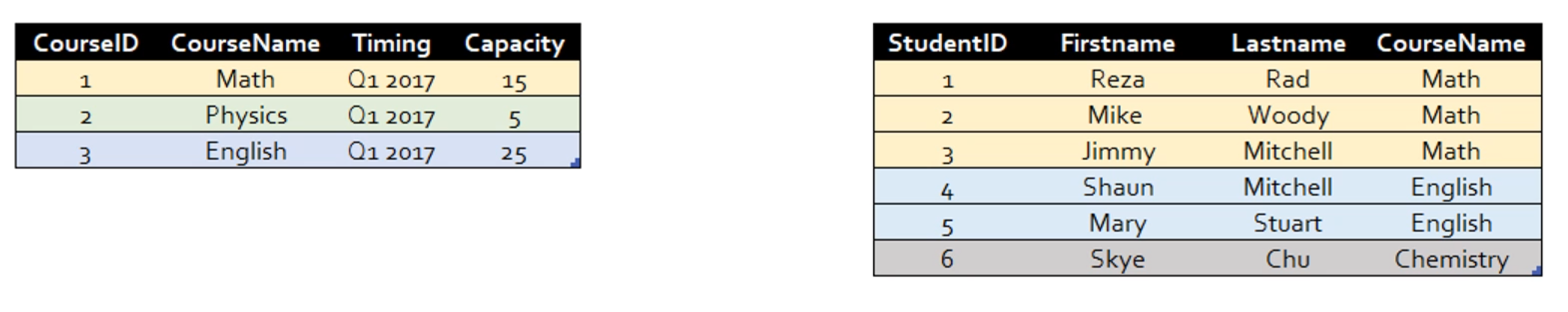
## Merge



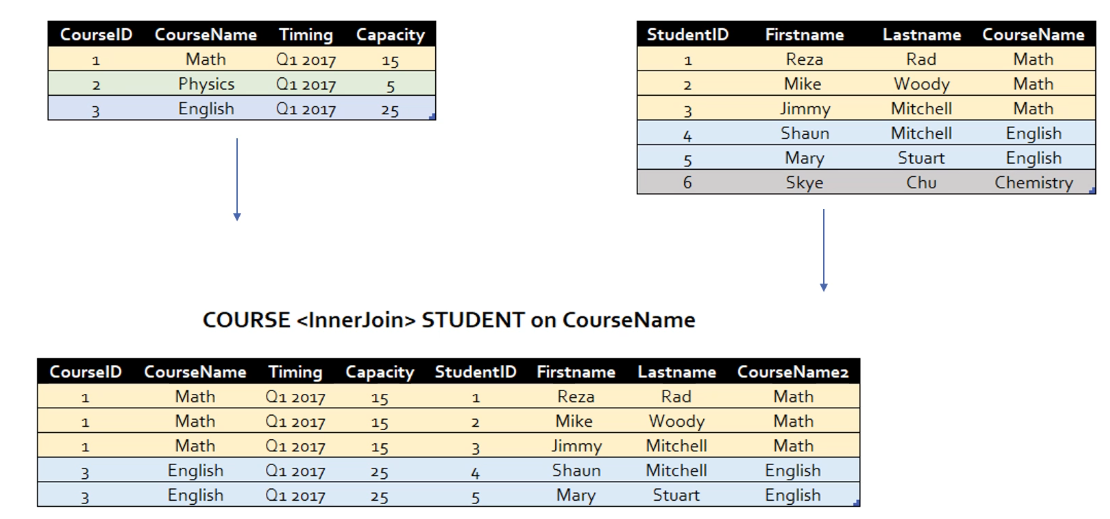
* We want to identify the type of merge. Different types of joins will result in different populated data.

## Example

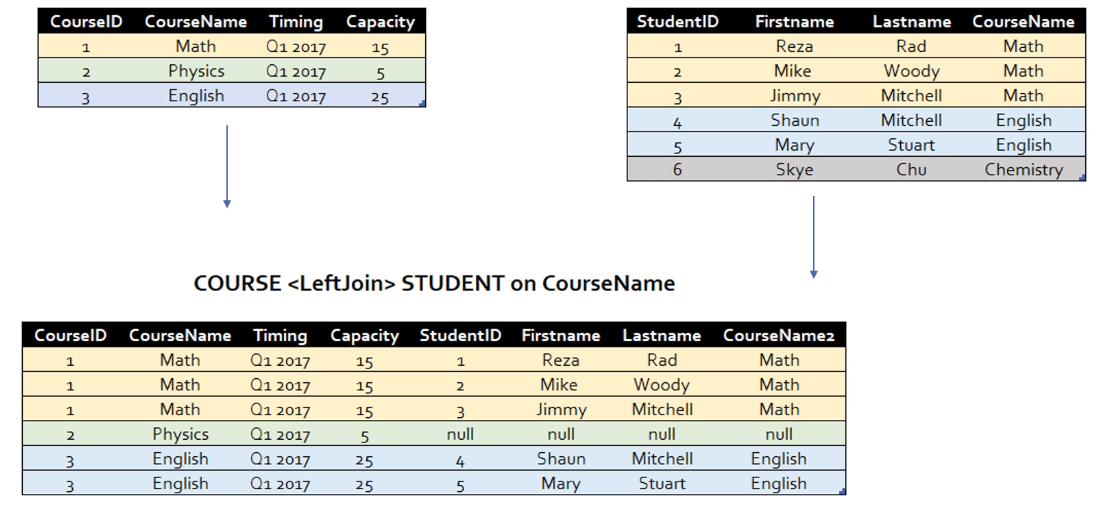
* Course vs Student table



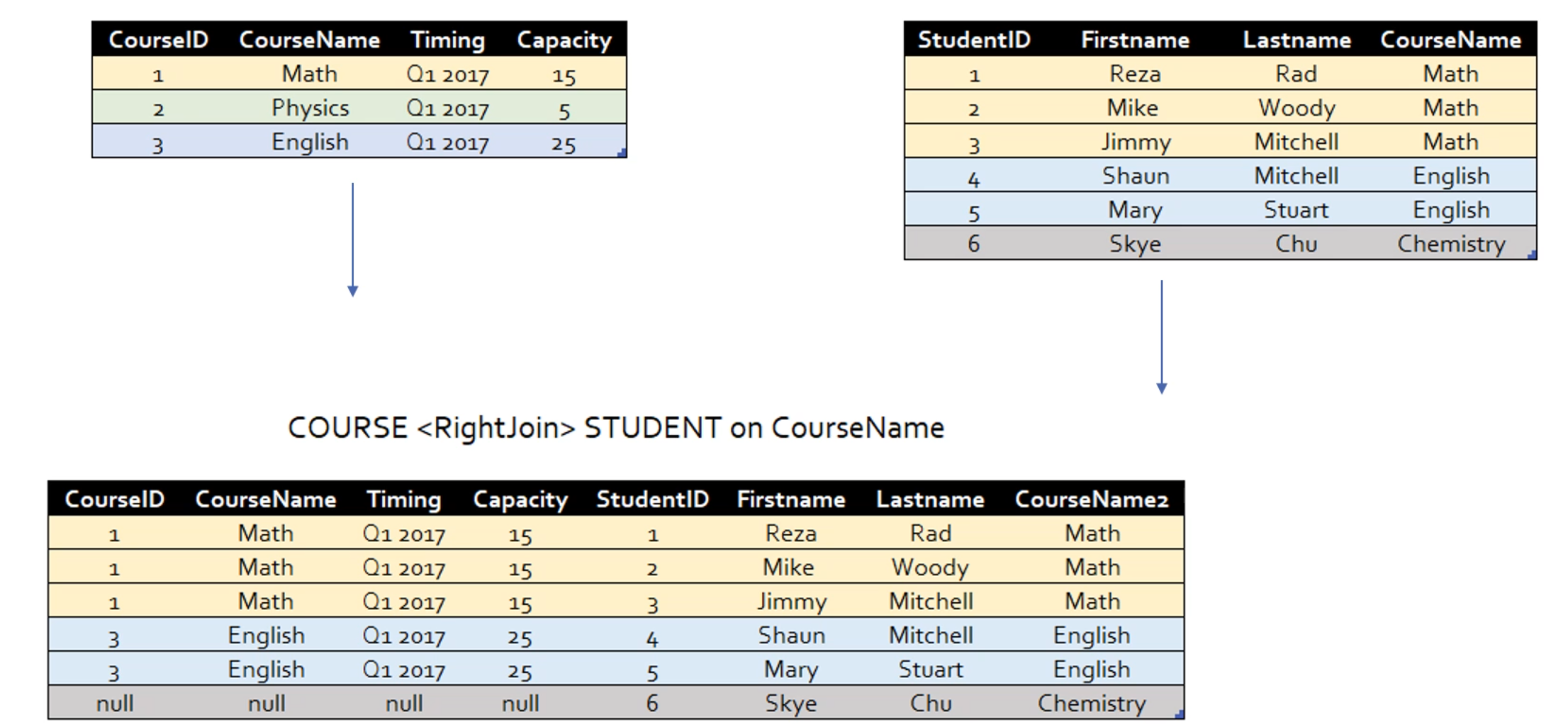
INNER JOIN



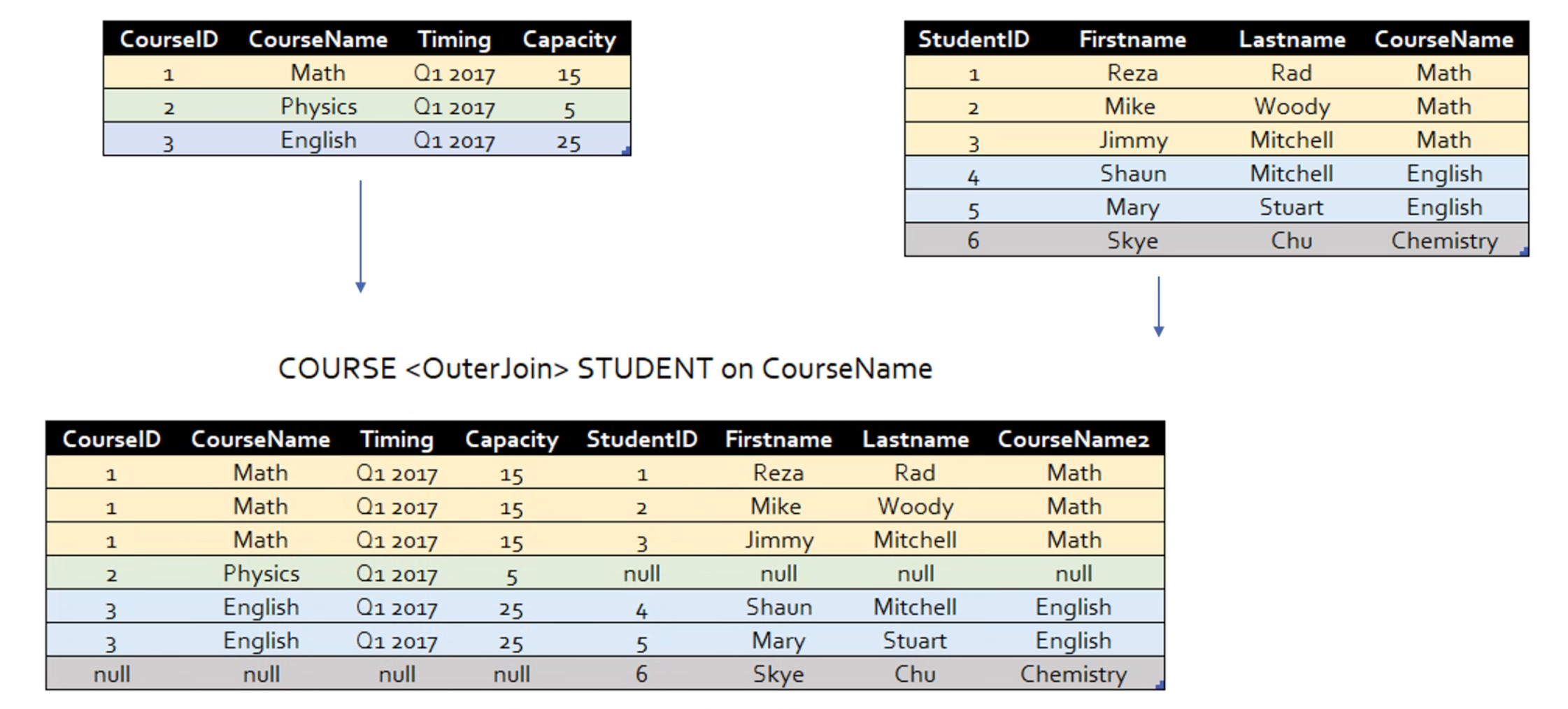
LEFT OUTER JOIN



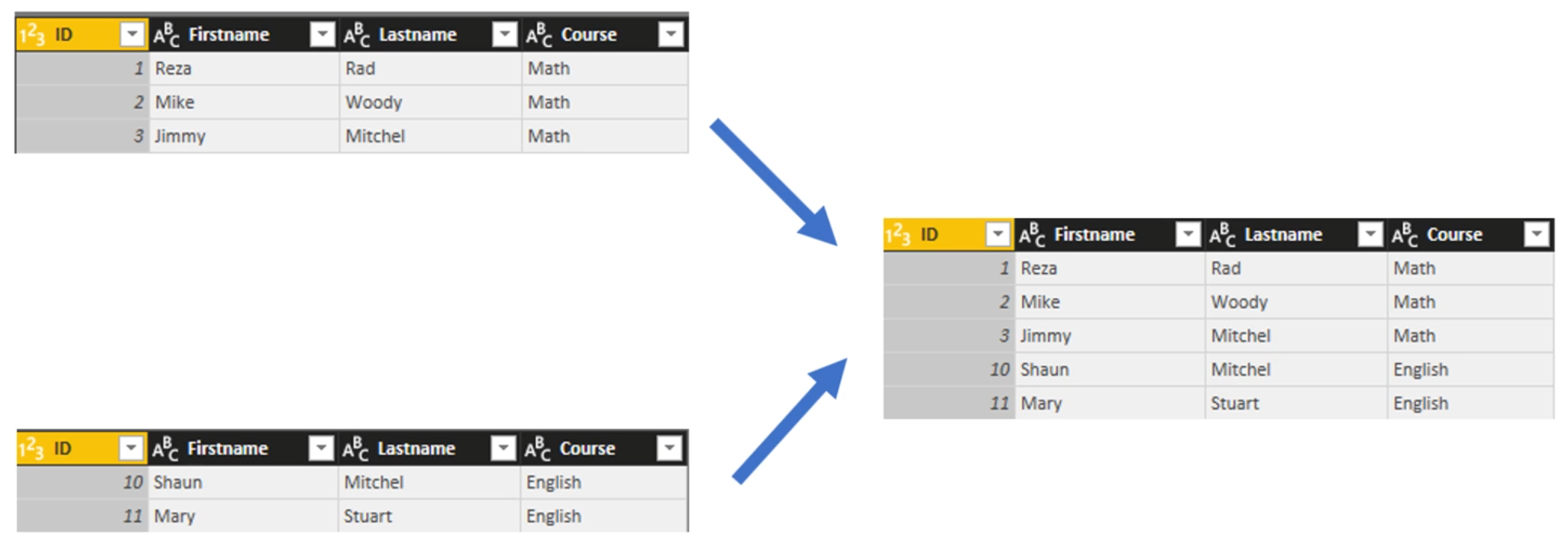
RIGHT OUTER JOIN



FULL OUTER JOIN



**APPEND**



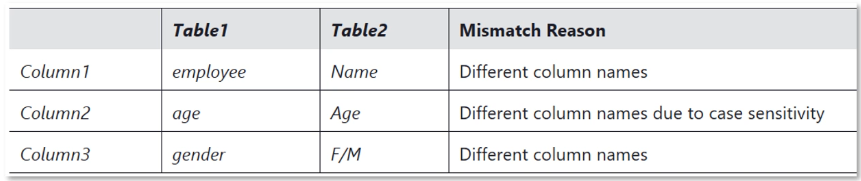
* Append will require the tables on being identical
* Appends tables with the same information with the same number of columns, same fields, named the same and they wont change

## What are Mismatched Tables?

Mismatched tables are tables that describe the same semantic entites and their facts but have mismatched column names.

For example, say that:

* Table 1 consists of the column names employee, age, and gender
* Table 2 consists of the column names Name, Age and F/M



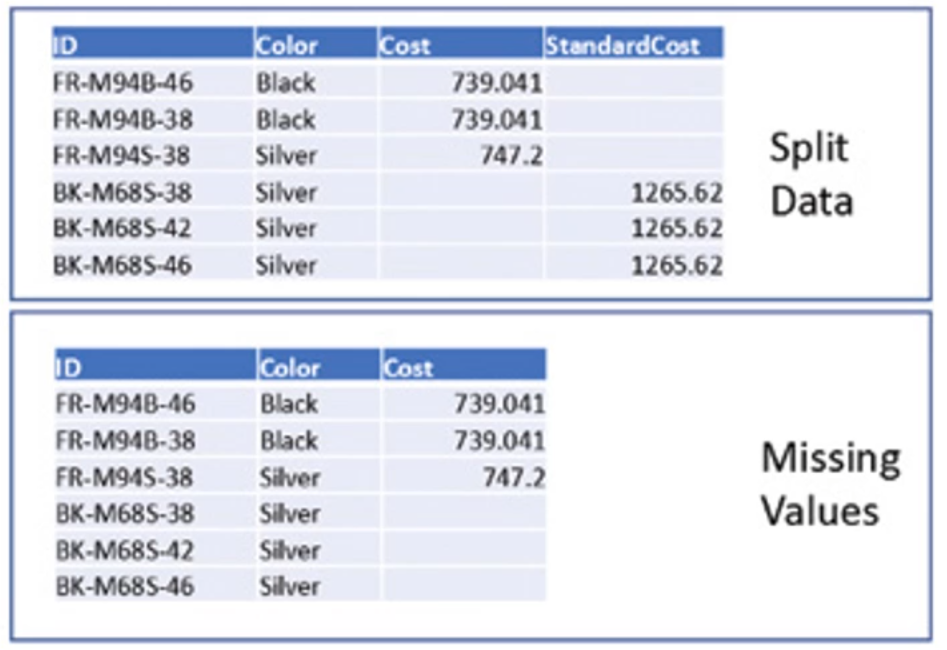
## Symptoms and Risks

Combining Mismatched Tables manifests two basic anomalous behaviours:

* Split Data
* Missing Values

The Power Query Editor generally creates a superset of data columns

* Append results contains all columns from Table2 and Table2, including common columns
* Data only appears in the column it was originally in



* If the columns or fields data types are different

## Resolving Mismatched Data

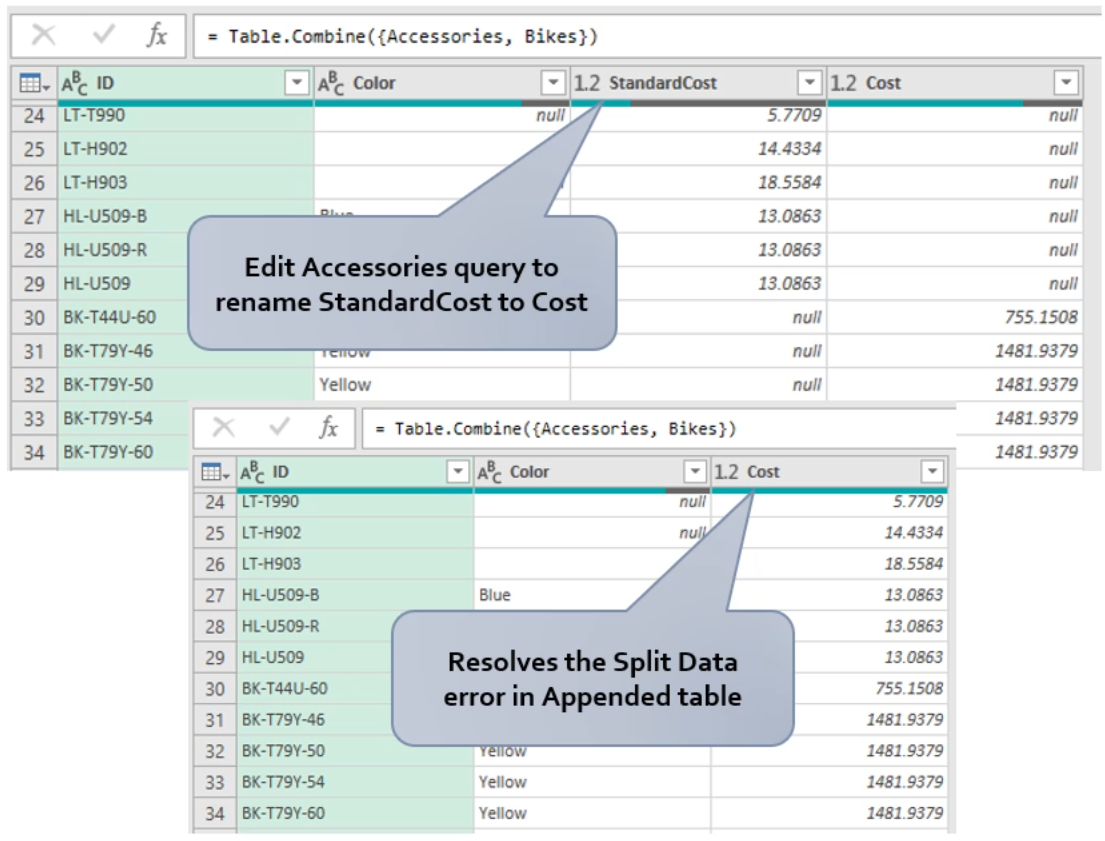
A Reactive Approach

* Edit the source queries to align Mismatched Tables column names

Using Append Queries as New allows for editing source queries

Appended query will automatically refresh to adjust to any source query changes

This is not very feasible for larger data models!



## Summary

Many methods exist to standardize column names for append operations. Which one you use depends on the level of detail you have in your tables, and your own commitment in terms of time and knowledge of M.

