

LESSON HANDOUT

Joining data in Excel

Joining data is the process of merging two or more tables together

Why

Joining data is an important function because:

- It enables the analysis of data derived from different sources
- Data must often be joined for the purpose of visualisation

Common fields

A join between data tables can only be achieved using a column which is common to all the data tables concerned. Think of the common column as the “glue” which allows the two tables to stick together.

- Common columns must be unique to one of the data tables concerned.
- Common columns are often Identification fields.
- The common columns must be in exactly the same format in both tables in order for the join to work.
- **Example:** the below tables share both the Customer ID and customer names field in common.
 - Without these common columns, a join between these two tables would not be possible.
 - The Customer ID field is unique to the customer table i.e. only one record of each Customer ID can exist in that table, whereas many Customer ID values can exist in the sale table - sale to many customers).
 - It is preferable to join these two tables with the Customer ID because any changes to the text of the Customer name (such as spelling) would result in an error.

Table 1 - Sales Table				
Ser	Customer ID	Sales	Customer Name	Segment
1	GH-14485	\$ 1,098	Gene Hale	Corporate
2	DV-13465	\$ 470	Dianna Vittorini	Consumer
3	CC-12100	\$ 627	Chad Cunningham	Consumer
4	CA-12265	\$ 392	Christina Anderson	Consumer
5	MB-18085	\$ 627	Mick Brown	Consumer
6	ME-17320	\$ 392	Maria Etezadi	Consumer
7	TC-21295	\$ 470	Toby Carlisle	Consumer
8	KM-16720	\$ 980	Kunst Miller	Consumer
9	CS-12490	\$ 235	Cindy Schnelling	Corporate
10	JB-15400	\$ 100	Jennifer Braxton	Corporate
11	KD-16270	\$ 241	Karen Daniels	Consumer
12	DW-13480	\$ 15	Dianna Wilson	Consumer

Table 2 - Customer Table			
Ser	Customer ID	Customer Name	Segment
1	GH-14485	Gene Hale	Corporate
2	DV-13465	Dianna Vittorini	Consumer
3	CC-12100	Chad Cunningham	Consumer
4	CA-12265	Christina Anderson	Consumer
5	MB-18085	Mick Brown	Consumer
6	ME-17320	Maria Etezadi	Consumer
7	TC-21295	Toby Carlisle	Consumer
8	KM-16720	Kunst Miller	Consumer
9	CS-12490	Cindy Schnelling	Corporate
10	JB-15400	Jennifer Braxton	Corporate
11	KD-16270	Karen Daniels	Consumer
12	DW-13480	Dianna Wilson	Consumer

Figure 1 - Example of 2 tables that can be joined

Excel JOIN functions

- **VLOOKUP** - Returns a value from a specified column from one table based upon a common column from another table
 - = VLOOKUP (Table 1 common field, Table 2 range, Table 2 column to return, exact or approximate match)
 - Disadvantages - will not survive additional columns being added to table 2 and cannot join values to the left of the common field in Table 2.
- **MATCH** - Returns the row number of a value in a range
 - = MATCH (value, range)
- **INDEX** - Returns the value of at a specified row of a range
 - = INDEX (range, row)
- **INDEX + MATCH**. Combine the two formulas by using **MATCH** to find the common column value in Table 2 and **INDEX** to look up that value.
 - = INDEX (Table 2 range, MATCH(table 1 value, table 2 range))