

Solution to Exercise 18.C

We have added some extra values to the data to make the points covered here clearer.

1. We have not included calculated values in the tables. They can be calculated as required.

SalesOrder

order Date	order No	customer Code	customer Name	customer Town	customer OrderRef	productCode	productDescription	quantity	unit Price
25/08/1999	37291	CE102	Central Stores	Lytham St Annes	R20716	12-75	Sandwich spread 24x250g	3	18.00
						09-103	Brown sauce 30x500g	10	24.60
26/08/1999	37292	CE102	Central Stores	Lytham St Annes	R20721	01-10	Evans Original Pickle 24x250g	8	30.00
26/08/1999	37299	BE005	Beckenhams	Stourport	6717	12-75	Sandwich spread 24x250g	5	18.00
						09-100	Tomato ketchup 30x500g	5	23.00

2. Each line in the sales order is uniquely identified by the Order No. However, these lines contain column -row intersections that contain multiple values. The first step is to sort out the multiple values or 'repeating groups' by creating a separate row for each one and copying the single values into these rows. The table below is now in **First Normal Form**.

SalesOrder-1

order Date	order No	customer Code	customer Name	customer Town	customer OrderRef	productCode	productDescription	quantity	unit Price
25/08/1999	37291	CE102	Central Stores	Lytham St Annes	R20716	12-75	Sandwich spread 24x250g	3	18.00
25/08/1999	37291	CE102	Central Stores	Lytham St Annes	R20716	09-103	Brown sauce 30x500g	10	24.60
26/08/1999	37293	CE102	Central Stores	Lytham St Annes	R20721	01-10	Evans Original Pickle 24x250g	8	30.00

							24x250g		
26/08/1999	37299	BE005	Beckenhams	Stourport	6717	12-75	Sandwich spread 24x250g	5	18.00
26/08/1999	37299	BE005	Beckenhams	Stourport	6717	09-100	Tomato ketchup 30x500g	5	23.00

Note. We are assuming that the product descriptions including the number and size of the packages are atomic values, and cannot be broken down further.

3. The table above contains redundant data that we must now eliminate. The next step is to identify the columns that uniquely identify each row. Order No. is no longer a unique identifier, as there is more than one row with the same Order No. Each row is uniquely identified by a combination of Order No. and Product Code.

4. We now identify which columns are not dependent on the whole key (Order No. and Product Code) but on only part of it. Have a look at Section 18.5.2 to refresh your understanding of dependency.

Product Description and Unit Price are dependent on the Product Code, but not on the Order No. They do not change for different orders. So they should be moved into a separate table with Product Code.

The Order Date, Customer Code, Customer Name, Customer Town and Customer Order Ref. are only dependent on the Order No. They do not change for different values of Product Code, so they will be removed into a separate table as well.

This gives us the tables below in **Second Normal Form**.

SalesOrder-2

order Date	order No	customer Code	customer Name	customer Town	customer OrderRef
25/08/1999	37291	CE102	Central Stores	Lytham St Annes	R20716
26/08/1999	37293	CE102	Central Stores	Lytham St Annes	R20721
26/08/1999	37299	BE005	Beckenhams	Stourport	6717

SalesOrderLine

order No	productCode	quantity
----------	-------------	----------

37291	12-75	3
37291	09-103	10
37293	01-10	8
37299	12-75	5
37299	09-100	5

Product

productCode	productDescription	unit Price
12-75	Sandwich spread 24x250g	18.00
09-103	Brown sauce 30x500g	24.60
01-10	Evans Original Pickle 24x250g	30.00
09-100	Tomato ketchup 30x500g	23.00

5. Each table now has a different unique identifier. For SalesOrder-2 it is Order No. For SalesOrderLine it is the combination of Order No. and Product Code. For Product it is Product Code.

We now review the tables to find out if there are any columns whose values are not dependent on the unique identifier in each table. These are 'non-key dependencies'. In the SalesOrder-2 table, the customer details (Customer Name and Customer Town) are not dependent on the Order No. (they do not change for each different order for that customer). However, the Customer Order Ref. does depend on the Order No. It is the customer's way of uniquely identifying each order in their own system.

We remove the columns that are dependent on a value other than the key (unique identifier) of this table.

SalesOrder-3

order Date	order No	customer Code	customer OrderRef
------------	----------	---------------	-------------------

25/08/1999	37291	CE102	R20716
26/08/1999	37293	CE102	R20721
26/08/1999	37299	BE005	6717

Customer

customer Code	customer Name	customer Town
CE102	Central Stores	Lytham St Annes
BE005	Beckenhams	Stourport

SalesOrderLine

order No	productCode	quantity
37291	12-75	3
37291	09-103	10
37293	01-10	8
37299	12-75	5
37299	09-100	5

Product

productCode	productDescription	unit Price
12-75	Sandwich spread 24x250g	18.00
09-103	Brown sauce 30x500g	24.60

01-10	Evans Original Pickle 24x250g	30.00
09-100	Tomato ketchup 30x500g	23.00

6. These tables are now in **Third Normal Form**.

Note. Although in the text we have referred to these columns by names such as Product Code, in the tables we have conformed to our naming convention that all attributes begin with a lower case letter and words are concatenated together with a capital letter at the start of any subsequent words. We have also assumed that tables will become classes, and have applied our class naming conventions to them. However, we would now remove the **-3** from the **SalesOrder-3** table, to give the class name **SalesOrder**.