QuickFill®
Application Notes

Sales Analysis Using QuickFill and PivotTables

or, Unscrambling Rubik's Cube

In the early 1980s there was a popular puzzle called a Rubik's Cube, which looked much like a child's block except that any face could be rotated. Of course as soon as you had turned the six faces a few times, the colors became scrambled. The object of the puzzle was to rotate everything back again so that each face of the cube was a uniform color—this, of course, was much harder than it looked. In this article I will attempt to explain how to unscramble your sales data to come up with a useful presentation. The tools we are going to use are QuickFill's ODBC driver and Microsoft Excel's PivotTables.

ODBC is one of those innumerable computer acronyms that techno geeks like to bandy about. It stands for *Open Database Connectivity*, but what you really need to know is that it is the means to access the information in your QuickFill database using many of the standard desktop applications.

A PivotTable is Microsoft's term for a cross tabulation. Suppose you have a list of all of your subscriptions and their current status, together with the subscriber's country, state and zip code, and the year and month they first subscribed. Your first thought might be to create a simple two-dimensional table with the states as the row headings and the years as the column headings. But that only satisfies you for a little while— now you want to see a breakout by state, year, and month. This requires a three-dimensional table, so you start thinking about a cube floating in midair with glowing numbers on each face. On one face of the cube you see subscriptions by state and year. On another face you see subscriptions by state and month, and on the third face you see subscriptions by year and month. This futuristic vision is at least conceivable—but what if you want to see a breakout by four items, say, state, year, month, and status? Unless you are a mathematician, cubes with more than three dimensions can make your head hurt. That is where a PivotTable comes in—it stores multidimensional data and rotates (pivots) it so that the face you are interested in is displayed on your two-dimensional computer screen.

Using Microsoft Excel, the basic procedure is to first retrieve a set of data from your QuickFill database. This data set may contain dates, zip codes, statuses, prices, payment amounts, and other similar information from the customer, subscription, and order records in the database. After retrieving the data, Excel stores it invisibly in your spreadsheet and lets you choose which combinations of data you want to appear in your table. You do this by dragging buttons from a field list to the row, column, and grid area of the table. Excel then automatically displays a cross tabulation of your chosen data fields.

For example, you might initially choose to display order prices by state and month with the states listed down the left side and the months across the top. In the body of the table, Excel shows you the total order prices for each state and month, thus giving you an idea of how your gross revenue is distributed. But then you decide that instead of the prices you would like to see the total payments received. Simply drag the Price field off the table and replace it with the Payment field. Excel automatically recalculates the table for you. It

QuickFill Application Notes (ISSN 1047-1030) is published by CWC Software, Inc., 150 Grossman Drive, Suite 201, Braintree, Massachusetts 02184. Address inquiries concerning this publication or Quickfill subscription fulfillment software to CWC Software, Inc., 150 Grossman Drive, Suite 201, Braintree, Massachusetts 02184. Telephone (800) 762-7702. Postmaster: Send address changes to *QuickFill Application Notes*, 150 Grossman Drive, Suite 201, Braintree, Massachusetts 02184.

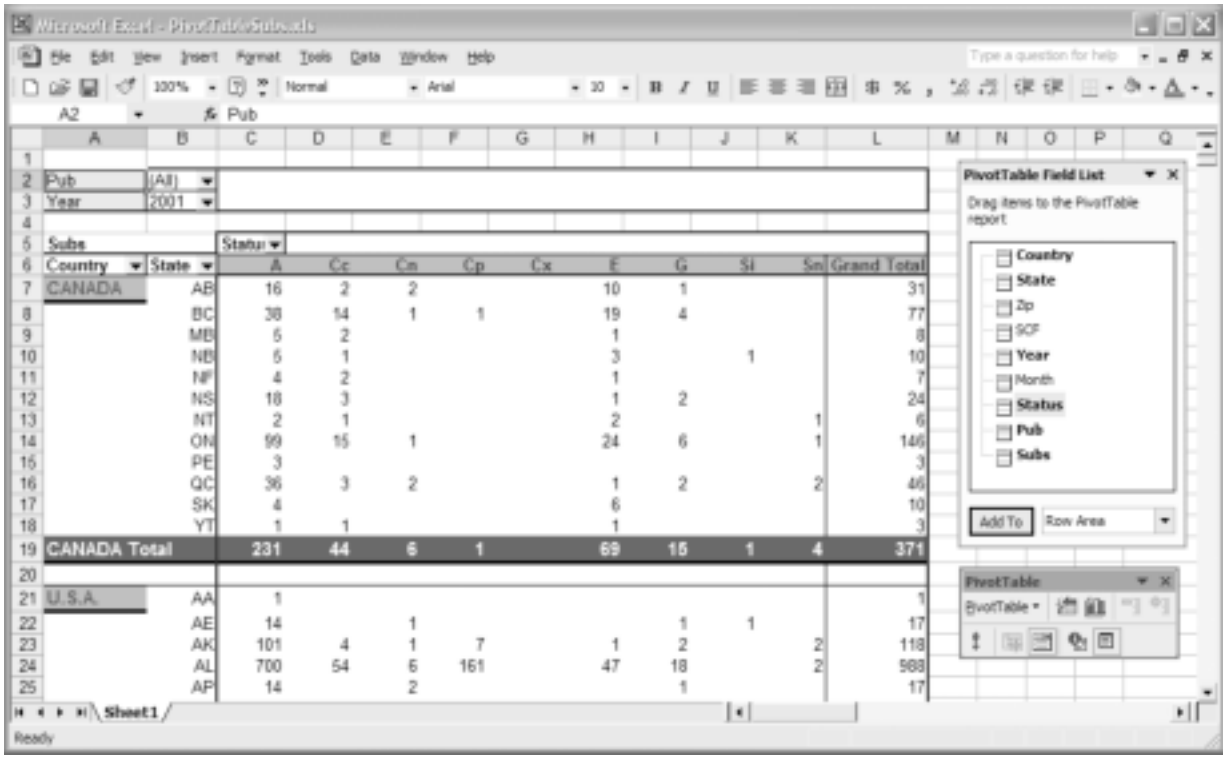

Figure 1

can do this reasonably quickly because it already has the data it needs stored in the spreadsheet and doesn't have to make another trip back to the QuickFill database to fetch the data.

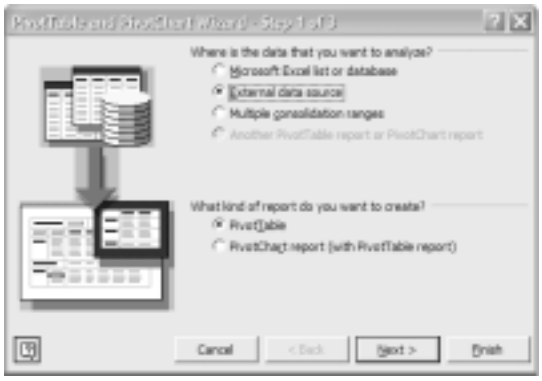

Figure 2

Let's try to construct a PivotTable like the one shown in Figure 1 using the data in your QuickFill database. In this example, Excel is displaying the number of subscriptions for the year 2001, broken down by country, state, and status. Many other combinations can be displayed simply by dragging fields from the field list on the right side of the screen.

The screen shots that follow are all from Excel 2002 (from Office XP), but almost everything I describe will also work in Excel 2000. I will highlight the places where they differ significantly.

Step 1: Starting with a blank worksheet, choose "PivotTable and PivotChart Report" from the Data menu. This brings up the PivotTable Wizard. (See Figure 2.)

Step 2: Choose "External data source" and "PivotTable" on the first screen of the wizard, then click the Next button. This brings up the second screen of the wizard, (see Figure 3) where it tells you that no data fields have been retrieved. Click the Get Data button to remedy that. This will start Microsoft Query, an adjunct to Microsoft Excel.

Note: If you have never used Microsoft Query before, the "install on demand" feature of Microsoft Office may kick in here and prompt you to insert your Microsoft Office CD.

Step 3: The first screen you see in Microsoft Query will be one labeled "Choose Data Source." Every QuickFill database that has ever been

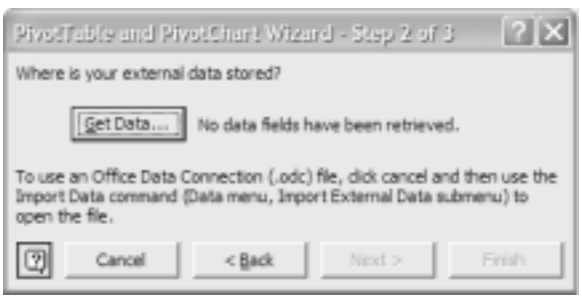

Figure 3

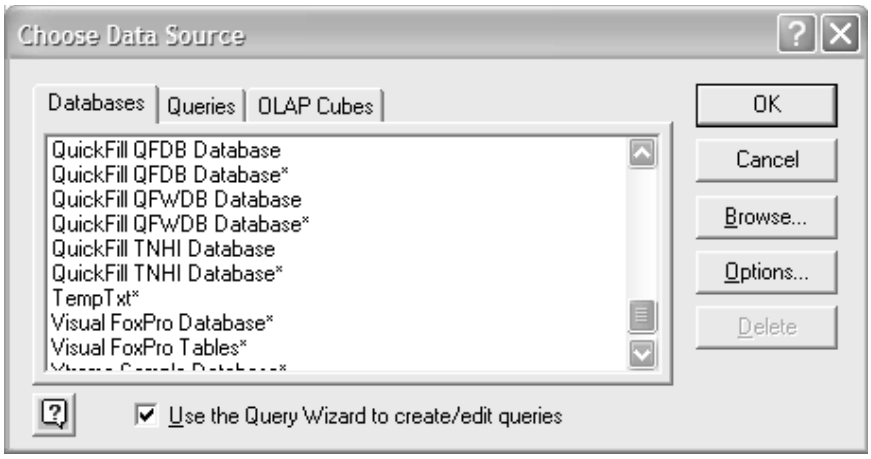

Figure 4

opened on your computer should be listed, so scroll down and select the one you are interested in. (See Figure 4.)

If there are duplicate entries, some with asterisks and some without, as in Figure 4, you can select either entry—it makes no difference. The duplicate entries are "file" data sources and "user" data sources. QuickFill creates both kinds because some versions of Microsoft Access or Excel show just one kind or the other. Recent versions of Access show you both.

Before you click OK, clear the check box on the bottom that is labeled "Use the Query Wizard to create/edit queries." Since we need to retrieve data from two tables (customers and subscriptions), the Wizard doesn't have quite enough magical powers for our purposes. Click OK and the main screen of Microsoft Query will appear, as shown in Figure 5.

Step 4: The Add Tables window of Microsoft Query lists all of the data tables in the QuickFill database, each identified by a three-character abbreviation. Think of a data table as a giant Excel grid where each row of the grid represents one record and each column represents one field in the record. For example, the customer table contains one row for each customer in the database and columns for each of the customer fields, such as customer number, last name, first name, and so on. Some of the more commonly used tables are:

PUB	Publications
CUS	Bill-to customers
SUB	Subscriptions
SHP	Ship-to customers
ORD	Orders

You can find a complete list of the tables in QuickFill by pressing F1 to open the Help window and looking up ODBC in the index. In our example, the data we want is in the customer table (CUS) and the subscription table (SUB). Select each of these two tables and click the Add button for each one and finally click the Close button.

Step 5: You should now have your CUS and SUB tables displayed as shown in Figure 6. Next, you need to "Join" your tables. Starting from the CUS_DBACUS line in the CUS table, hold the mouse button down and drag it over to the right until it is over the SUB_ DBACUS line in the SUB table, then let go. The result should be a line linking the CUS_DBACUS field and the SUB_DBACUS field, as shown above.

The purpose of joining the two tables is to let Microsoft Query know how to tell which subscription record belongs to each customer record. The fields in each record that contain "DBA" in their names are database address fields. Every subscription record contains a field that lists the database address of its customer record. Perhaps you can see the pattern in the field names now:

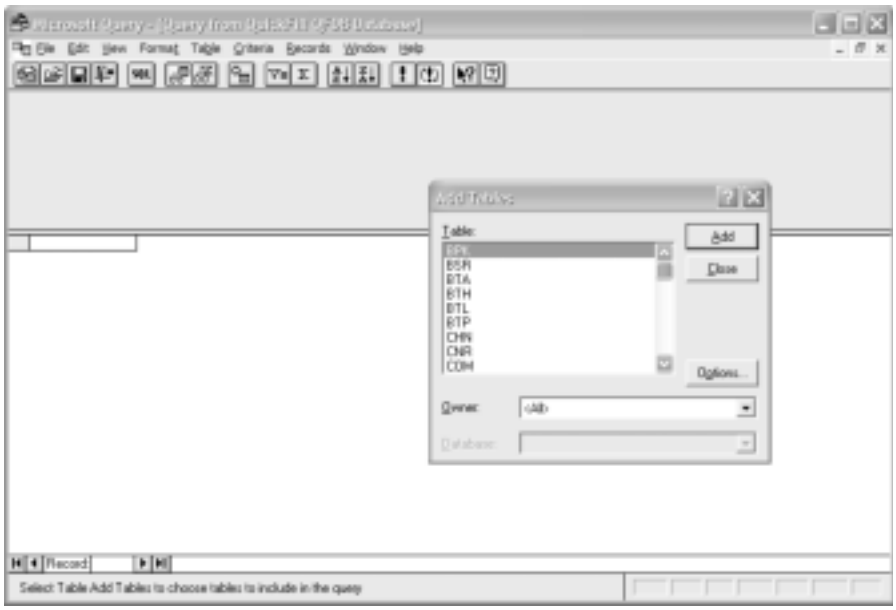

Figure 5

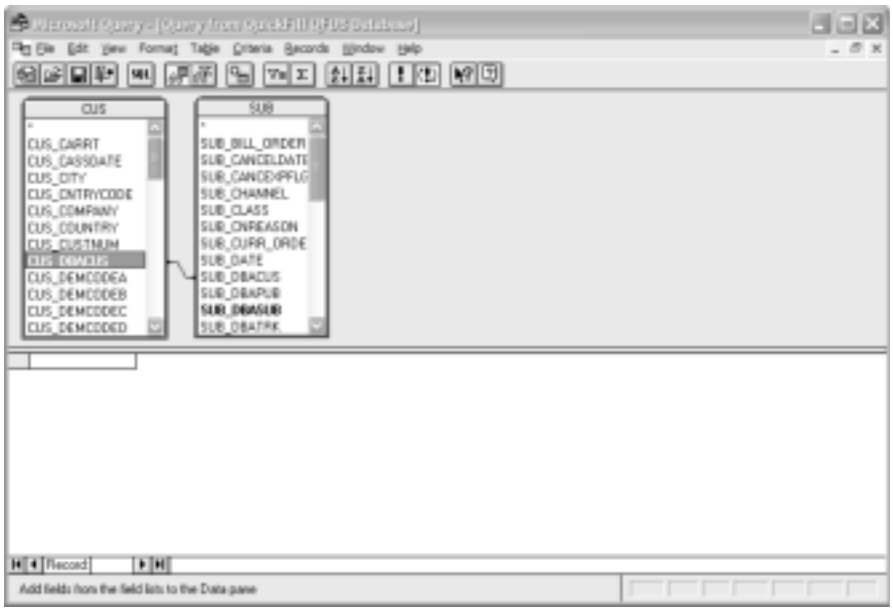

Figure 6

SUB_DBACUS breaks down into SUB (all subscription fields begin with SUB), DBA (for database address), and CUS (for the customer record). If you want to impress your boss tell him or her that the CUS_DBACUS field is the "primary key" of the CUS table and the SUB_DBACUS field is a "foreign key" that links the CUS table to the SUB table.

The best joins for the most commonly used tables are these:

PUB to SUB	Join PUB_DBAPUB to SUB_DBAPUB
SUB to ORD	Join SUB_DBASUB to ORD_DBASUB
SUB to SHP	Join SUB_DBASUB to SHP_DBASUB
CUS to SUB	Join CUS_DBACUS to SUB_DBACUS

Step 6: Now we choose the fields that we want from these two tables. These are the fields that we want and the names by which they are known in the QuickFill database:

Country	CUS_COUNTRY
State	CUS_STATE
Zip code	CUS_ZIP
Sectional center or SCF	CUS_ZIP
Year of original subscription	SUB_DATE
Month of original subscription	SUB_DATE
Status	SUB_STATUS
Publication code	SUB_PUB

Drag each of the fields listed above to the open area on the bottom of the screen as shown in Figure 7. Notice that some of the fields are listed twice. That is because we want different portions of those fields—the first five digits of the zip code, the first three digits of the zip code, and the year and month of the subscription date. In a minute we will fix the duplicate fields so that we get the portions we need. After you have dragged all of the fields to the bottom, it should look like the picture above. Don't worry if the order of the fields is different on your screen—the order doesn't matter. On the other hand, if you like being tidy you can drag the fields around on the screen until they match the order shown above.

Step 7: The next step is to assign nicer column headings to the database fields, so that we end up with just "Country" instead of "CUS_COUNTRY." Double-click each of the column headings on the bottom half of the screen in turn, and enter a new heading as shown in Figure 8. Use these column names:

CUS_COUNTRY	Country
CUS_STATE	State
CUS_ZIP	Zip
CUS_ZIP	SCF
SUB_DATE	Year
SUB_DATE	Month

Figure 7

SUB_STATUS	Status
SUB_PUB	Pub

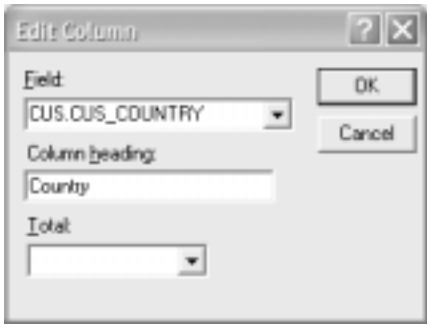

Figure 8

Step 8: Now we fix the zip code and date fields to extract the portions that we need. Double-click the Zip column heading and edit the "Field" box so that instead of CUS.CUS_ZIP it contains Left(CUS.CUS_ZIP,5). (See Figure 9.) This tells the ODBC driver

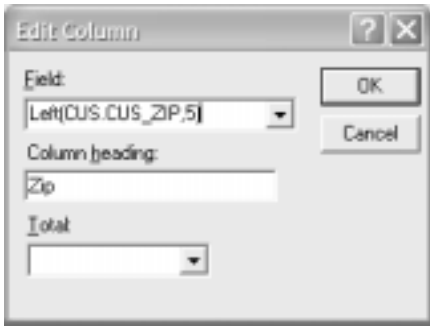

Figure 9

that we want just the first five digits of the zip code. Do this for each of these four fields:

Zip	Left(CUS.CUS_ZIP,5)
SCF	Left(CUS.CUS_ZIP,3)
Year	Year(SUB.SUB_DATE)
Month	Month(SUB.SUB_DATE)

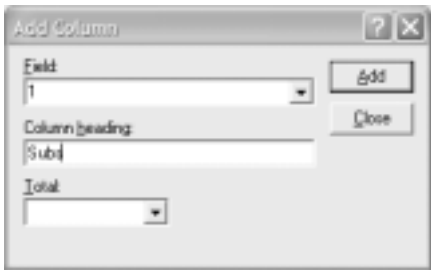

Figure 10

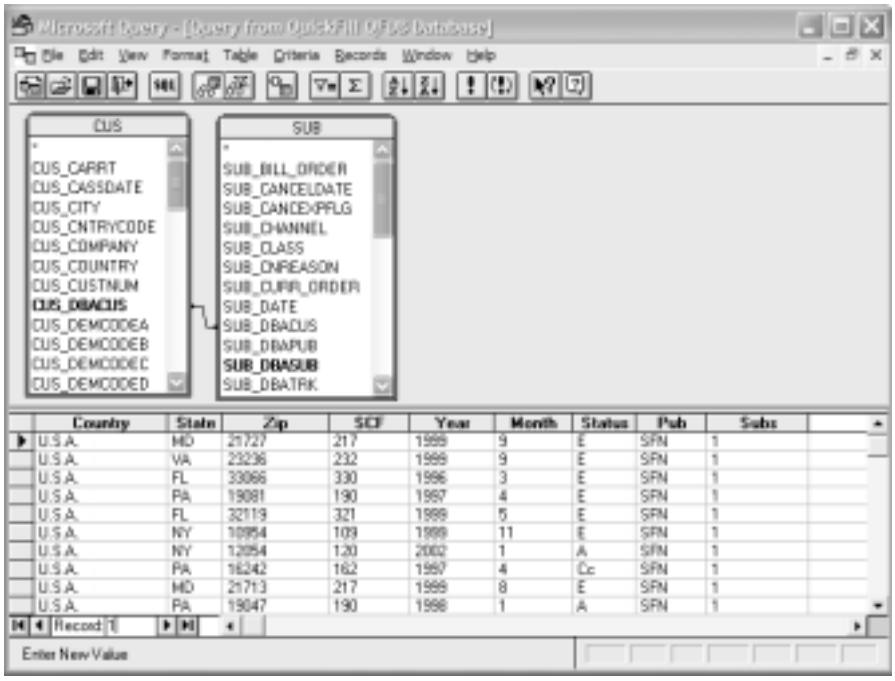

Figure 11

Step 9: By now, Microsoft Query should be showing you the first few rows of the data you have requested (if not, click the exclamation mark on the toolbar). Each row represents one subscription. In Excel we would like to count the number of rows with each possible combination of country, state, zip, year, month, and status, so we better give it something to count. To make it easy, we just add a value of 1 to each row of the table. Click in the empty column heading on the right side of the grid and then on the Records menu choose "Add column…." Enter a '1' in the Field box and "Subs" in the Column heading box, as shown in Figure 10. Click the Add button, the Close button, and then click the exclamation mark again. The result should look like Figure 11.

Step 10: Now that we have the data we need, it is time to go back to Excel and make our PivotTable. From the File menu choose "Return Data to Microsoft Excel." This takes you back to the PivotTable Wizard, where you can click the Next button to go to Step 3 of the Wizard. (See Figure 12.)

Step 11: On the Wizard's Step 3 screen, click the Layout button so we can make the initial arrangement of rows and columns in our PivotTable. For an example of the layout screen, see Figure 13.

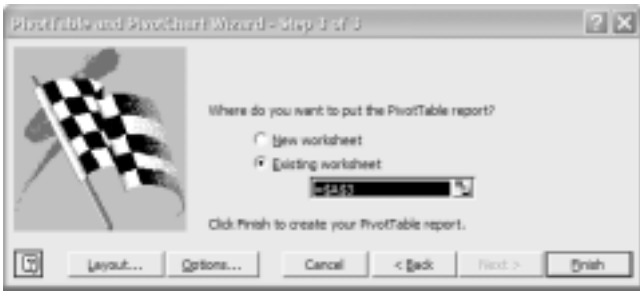

Figure 12

On the right side are buttons for each column of data that we obtained from the database. On the left side is a pictorial representation of the PivotTable. Drag the Subs button

Figure 13

over to the Data area, the State button over to the Row area, and the Status button to the Column area. Finally, drag the Pub and Country buttons to the Page area. The result should look like Figure 14.

Click the OK button to return to the Wizard. Finally, click the Finish button. Excel will retrieve all of the rows of data from the database and construct a PivotTable using the layout we specified on the screen above. This may take a while if your database is large. To assure yourself that something is really happening, you can watch the record counter that Excel displays on the lower left side of the screen. The resulting PivotTable should look something like Figure 15.

If you are using Excel 2000, things will look a little different. The PivotTable toolbar and the field list are combined into a single window. If you hover the mouse over the field list, you will see that the field names are really buttons that can be dragged over to the table.

At this point, I would suggest saving your spreadsheet. All of the data is invisibly stored in the spreadsheet while the PivotTable presents various views of that data. You don't need to go back to the QuickFill database unless you want to get a different set of data or more current data.

What we've got is a cross tabulation of subscriptions by state and status. Now the fun begins—we can play with the table to see different "faces" of our multidimensional cube. Try some of these things:

1. From the drop-down list next to the Country field on the top left of the screen, pick U.S.A. This will limit the table to subscriptions from U.S. subscribers.

2. Drag the Year field from the "PivotTable Field List" and drop it just below the Pub and Country fields. Now you can choose to display any publication, country, or year. Or you can choose all publications, countries, and all years.

3. Click the arrow next to the Status field. This will produce a drop-down list of all of the possible statuses. Uncheck the four cancellation statuses (Cc, Cn, Cp, and Cx) and the expiration status. Click OK and you will have a table that includes only active, graced, and suspended subscriptions.

Figure 14

4. Drag the Status field off the table and put it back in the field list. Now you have a single column table broken out by state.

5. Drag the Year field from the top left and put it in the column heading area. Now you have a two-dimensional table by state and year.

6. Drag the Month field from the field list and drop it just below the Year field. Now you have a three-dimensional table by state, year, and month. If the year and month end up in the wrong order, just drag and drop to rearrange them.

7. Right-click on the '1' representing the first month of the first year and choose "Group and Show Detail" and then "Group" from the context menu. Tell Excel to start at 1, end at 12, and group by 3. Click OK and you have consolidated the months into quarters.

In Excel 2000 the menu entry is named "Group and Outline" instead of "Group and Show Detail."

8. In the drop-down list for State, clear the checkmarks from all of the states except your home state. Drag the Zip field from the field list and drop it just to the right of the State field. Now you have a breakdown of subscriptions by zip code within your home state, arranged by year and quarter. That's a four-dimensional cube!

If you are using Excel 2000, this last step may result in the following message: "A field in your source data has more unique items than can be used in a PivotTable." Excel 2000 cannot process more than 8,000 unique values per field. In Excel 2002 this limit was raised to 32,500. See page 8 for

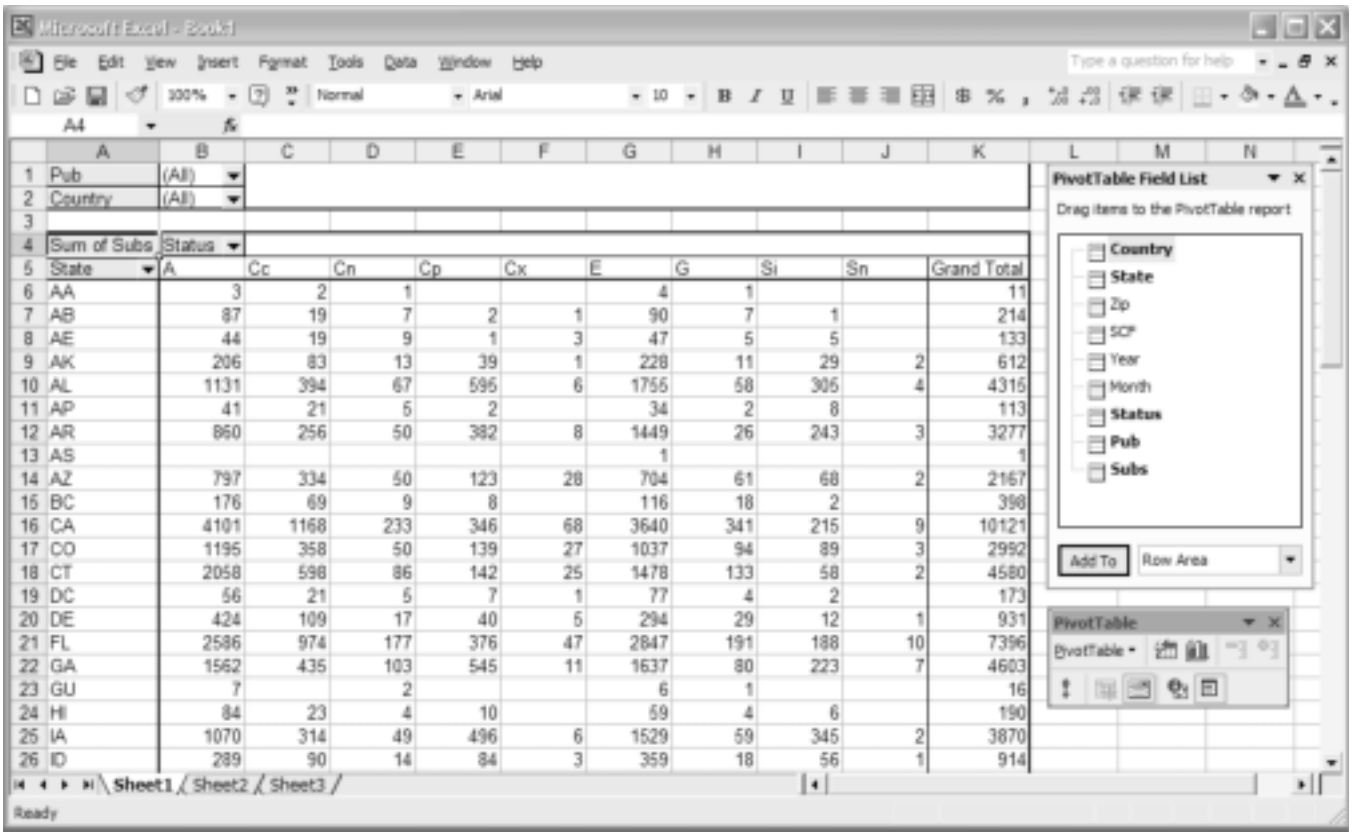

Figure 15

instructions on filtering your data to reduce the number of unique zip code values.

Now that we have seen a sample PivotTable, let's go back and look at some of those steps in more detail to see what else we might have done.

In Step 4 we selected the customer and subscription tables, whose abbreviations are CUS and SUB. Another commonly used table is the order table, which goes by the name of ORD. Every subscription has at least one order. Each time the subscription is renewed another order is added, so a subscription that has been active for several years will typically have one order record for each year. You can join the order table to the subscription table by linking the fields SUB_DBASUB in the SUB table to the ORD_DBASUB field in the ORD table.

Why would you want to add the order table to your query? One reason would be to get information about new and renewal orders received over a time period. If you analyze the subscription dates, you are only looking at the dates of the original orders. Somebody who subscribed 10 years ago and has renewed every year since wouldn't show up in a tally of subscriptions received during the last 12 months.

Another common reason for adding the order table to your query is to get financial data. The order record holds information about the price of the order and the amount paid (ORD_PRICE and ORD_FPAY). These are useful if you want to analyze the amount of money you received over a period of time or by geographic area. But there are a couple of things you should know before you use these data fields.

First, QuickFill stores all money amounts in pennies, not dollars. So if you are looking at the data using Excel, you will see a value of 2,000 where you expected to see a price of 20.00. QuickFill does this because computers can store whole numbers more accurately than decimal values.

Second, QuickFill stores some money fields as negative values. Price, tax, and shipping are all considered to be credit fields by the folks who wear green eyeshades, so they have negative values. The payment field is a debit field, so it will have a positive value.

In summary, if you want to analyze the payments received from orders, then you should add "ORD_FPAY/100" to your query. The /100 divides the value by 100, giving you dollars and cents instead of pennies. If you want to analyze the prices of orders,

then you should add "ORD_PRICE/-100" to your query. The minus sign changes the sign and the /100 converts the result to dollars and cents. Don't try and put the minus sign in front of the ORD_PRICE; you will get an SQL syntax error.

You may have also noticed another table that was mentioned in Step 4, the ship-to or SHP table. This table is very similar to the customer (CUS) table. It contains all the same fields but with a SHP prefix instead of a CUS prefix. For single subscriptions, where the shipping and billing addresses are the same, it makes no difference whether you get your data from the CUS or the SHP table; the result will be the same. For two-party subscriptions, the two addresses are usually different and you must choose whether you wish to analyze the data from the shipping addresses or the billing addresses. Since group subscriptions can have multiple shipping addresses, you will get a higher subscription count if you include the SHP table than if you include the CUS table. Also, be warned that if you are looking at prices or payments, then the totals will be overstated for group subscriptions if you include the SHP table in your query. Generally, it is best to use the CUS table instead of the SHP table unless you are specifically interested in the geographic areas that your publication is mailed to.

In Step 6 we chose the data fields that we wanted to include in our table. Here are some other commonly used fields that you can choose from:

SUB_TRACK	Tracking code
SUB_CHANNEL	Original channel code
SUB_LIST	List code
SUB_PLAN	Plan code
SUB_EXPISS	Expire issue number
SUB_EXPISSDATE	Expire issue date
ORD_ISSUES	Term
ORD_QUANTITY	Number of copies ordered
ORD_SERVICE	Service code
ORD_PREMIUM	Premium code
CUS_COMPANY	Company name

Filtering your data

Analyzing five-digit zip codes or company names can be tricky due to the large number of zip codes or company names that may exist in your database. Microsoft Excel may bog down if it has to process more than a few thousand distinct company names. To reduce the number of zip codes or company names in your data set, you can add some filter criteria in Microsoft Query. You should do this after Step 9 and before Step 10 by following this procedure. First, select "Criteria" on the View menu—this will display an additional grid in the center of the window. Second, choose "Add Criteria" from the Criteria menu. You will get a window that looks like Figure 16.

Figure 16

In the Field drop-down list, you should choose the field that you are going to filter on. In the Operator drop-down list, choose the type of comparison that should be performed (equals, does not equal, is greater than, etc.). In the Value field, enter the code value to be compared against. Finally, click the Add button and Microsoft Query will add the criteria to the criteria grid.

In the example above we are restricting our query to records involving customers in Massachusetts. This is an effective means of reducing the number of zip codes or company names to a number that Microsoft Excel can handle.

If you wish, you can add multiple criteria to your query to expand or further cut down on the volume of data that must be retrieved. You might want to select all of the states in a particular region of the country. To include additional states, just add their codes to the criteria grid underneath the first state. Alternatively, you might want to select just those subscriptions that are active. To do that, you would add an additional column to the criteria grid selecting the value 'A' in the Status field.

Any pruning down of the data set that you do in Microsoft Query will speed things up when you get back to Microsoft Excel. The flip side of the coin is that limiting the data set also limits your flexibility when rearranging your PivotTable in Excel.

For more information about using PivotTables in Microsoft Excel, I can recommend the book **Microsoft Excel Version 2002 Inside Out**, by Craig Stinson and Mark Dodge, published by Microsoft Press. Chapter 30 discusses PivotTables in great detail.