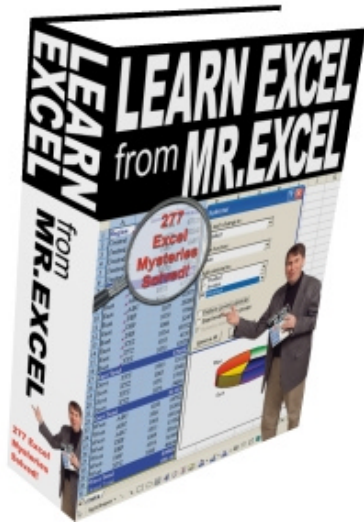




Free Preview – Week #17
Learn Excel from Mr Excel
Copyright 2005 Bill Jelen
All Rights Reserved

Encourage your friends to sign up at
<http://www.mrexcel.com/learn-excel.html>

This week - four topics that deal with imperfect data.
The first two topics introduce concepts that come in
handy when solving the problem in topic 3.

NUMBER EACH RECORD FOR A CUSTOMER, STARTING AT ONE FOR A NEW CUSTOMER

Part
III

Problem: You have a list of invoice data, as shown in Fig. 810. You wish to number the records in such a way that the first invoice number for Ford is one. The next Ford invoice is two, and so on. When you get to a new customer, you want to start over at one.

	A	B	C	D	E	F	G	H	I
1	Invoice	Customer	Date	Region	Product	Quantity	Revenue	COGS	Profit
2	1010	Ford	1/1/04	East	XYZ	1000	22810.00	10220	12590
3	1011	Verizon	1/2/04	Central	DEF	100	2257.00	984	1273
4	1012	Verizon	1/2/04	East	ABC	500	10245.00	4235	6010
5	1013	Ainsworth	1/3/04	Central	XYZ	500	11240.00	5110	6130
6	1014	Ainsworth	1/4/04	Central	XYZ	400	9204.00	4088	5116
7	1015	Gildan Activew	1/4/04	East	DEF	800	18552.00	7872	10680
8	1016	Texaco	1/4/04	East	XYZ	400	9152.00	4088	5064
9	1017	IBM	1/5/04	Central	ABC	400	6860.00	3388	3472
10	1018	General Motors	1/7/04	East	ABC	400	8456.00	3388	5068
11	1019	State Farm	1/7/04	East	DEF	1000	21730.00	9840	11890
12	1020	Texaco	1/7/04	West	XYZ	600	13806.00	6132	7674
13	1021	General Motors	1/9/04	Central	ABC	800	16416.00	6776	9640
14	1022	HP	1/9/04	East	XYZ	900	21015.00	9198	11817
15	1023	Ainsworth	1/10/04	East	XYZ	900	21465.00	9198	12267

Fig. 810

Strategy: Sort the data by Customer. Insert a new temporary column A and add a heading called “Rec #”, as shown in Fig. 811.

	A	B	C	D	E
1	Rec #	Invoice	Customer	Date	Region
2		1013	Ainsworth	1/3/04	Central
3		1014	Ainsworth	1/4/04	Central
4		1023	Ainsworth	1/10/04	East
5		1025	Ainsworth	1/12/04	West

Fig. 811

The formula in A2 is =IF(C2=C1,1+A1,1). In plain language, this formula says, “If the customer in C is equal to the customer above me, then add 1 to the cell above me. Otherwise, start at 1.”

Result: Each group of customer invoices is numbered from 1 to N, as shown in Fig. 812.

	A54				
	A	B	C	D	E
1	Rec #	Invoice	Customer	Date	Region
52	51	1562	Ainsworth	12/24/04	West
53	52	1566	Ainsworth	12/26/04	Central
54	1	1143	Air Canada	3/31/04	West
55	2	1169	Air Canada	4/19/04	Central
56	3	1368	Air Canada	8/24/04	West
57	4	1438	Air Canada	10/7/04	East
58	1	1165	Bell Canada	4/13/04	East
59	2	1271	Bell Canada	6/24/04	Central
60	3	1382	Bell Canada	8/31/04	West
61	4	1479	Bell Canada	11/4/04	West
62	1	1086	Chevron	2/26/04	Central

Fig. 812

As shown in Fig. 813, copy the formulas in column A and Paste Special – Values to change to numbers so that you can re-sort the data by Invoice Number.

	A	B	C	D
1	Rec #	Invoice	Customer	Date
2	1	1010	Ford	1/1/04
3	1	1011	Verizon	1/2/04
4	2	1012	Verizon	1/2/04
5	1	1013	Ainsworth	1/3/04
6	2	1014	Ainsworth	1/4/04
7	1	1015	Gildan Activewear	1/4/04
8	1	1016	Texaco	1/4/04
9	1	1017	IBM	1/5/04

Fig. 813

Part
III

Summary: The IF function is perfect for this task of comparing the current record to the record above.

Functions Discussed: =IF()

ADD A GROUP NUMBER TO EACH SET OF RECORDS WITH A UNIQUE CUSTOMER NUMBER

Problem: You have a list of invoice data. You wish to number the records in such a way that the invoices for the first customer all have a group number of 1. The invoices for the next customer all will have a group number of 2.

Strategy: Sort the data by customer. Add a new column A, with the heading of Group. In cell A2, enter the number 1 for Group #1. In cell A3, enter the formula that will be used for the rest of the records:

=IF(C3=C2,A2,1+A2)

DEAL WITH DATA WHERE EACH RECORD TAKES FIVE PHYSICAL ROWS

Problem: Sometime back, in the days of COBOL, a programmer was dealing with the constraints of the physical width of a page. The programmer built a report where each record actually took up five lines of the report, as shown in Fig. 815. You now want to analyze this data in Excel.

	A3		fx =IF(LEFT(C3,4)="ACCT",1+A2,A2)				
	A	B	C	D	E	F	G
1	Group	Seq	Text				
2		1	ACCT: 12345		INVOICE: 1010		
3		1	INVOICE TOTAL: \$125.00				
4		1	ABC CO				
5		1	123 S. MAIN STREET				
6		1	SALEM OH 44460				
7		1	-----				
8		2	ACCT: 23456		INVOICE: 1011		
9		2	INVOICE TOTAL: \$175.00				
10		2	XYZ INC.				
11		2	456 N. BROADWAY				
12		2	SALEM OR 98754				
13		2	-----				
14		3	ACCT: 34567		INVOICE: 1012		
15		3	INVOICE TOTAL: \$225.00				
16		3	HANEY & ASSOCIATES				
17		3	789 LUNDY LANE				
18		3	SALEM MA 12345				

**Part
III**

Fig. 815

Strategy: Your goal is to get the data back into one row per record. This process is possible. The process involves adding two new columns, one called Group number and one called Sequence.

- 1) First, add a new Row 1. Insert two new columns A and B. Add headings as shown in Fig. 816 in A1:C1. The headings should be Group, Sequence, and Text.

	A	B	C
1	Group	Seq	Text
2			ACCT: 12345
3			INVOICE TOTAL: \$125.00
4			ABC CO

Fig. 816

- 2) In column A, assign a Group number to each logical record.

One way to do this is to check to see if the first four characters of column C are "ACCT". If this is true, add one to the group number. In A2, enter the number 1. In A3, enter the formula =IF(LEFT(C3,4)="ACCT",1+A2,A2). Copy it down to all of the rows. This will neatly assign each logical group of records a group number, as shown in Fig. 817.

	A3		=IF(LEFT(C3,4)="ACCT",1+A2,A2)				
	A	B	C	D	E	F	G
1	Group	Seq	Text				
2	1		ACCT: 12345		INVOICE: 1010		
3	1		INVOICE TOTAL: \$125.00				
4	1		ABC CO				
5	1		123 S. MAIN STREET				
6	1		SALEM OH 44460				
7	1		-----				
8	2		ACCT: 23456		INVOICE: 1011		
9	2		INVOICE TOTAL: \$175.00				
10	2		XYZ INC.				
11	2		456 N. BROADWAY				
12	2		SALEM OR 98754				
13	2		-----				
14	3		ACCT: 34567		INVOICE: 1012		
15	3		INVOICE TOTAL: \$225.00				
16	3		HANEY & ASSOCIATES				
17	3		789 LUNDY LANE				
18	3		SALEM MA 12345				

Fig. 817

3) Next, design a formula for a sequence number.

In cell B2, enter the formula `=IF(A2=A1,B1+1,1)`. Copy this down. This formula will number each record in the group, as shown in Fig. 818. It should ensure that all of the Account numbers are on a Sequence 1 record.

	B2		fx =IF(A2=A1,B1+1,1)			
	A	B	C	D	E	
1	Group	Seq	Text			
2	1	1	ACCT: 12345	INVOIC		
3	1	2	INVOICE TOTAL:	\$125.0		
4	1	3	ABC CO			
5	1	4	123 S. MAIN STREET			
6	1	5	SALEM OH 44460			
7	1	6	-----			
8	2	1	ACCT: 23456	INVOIC		
9	2	2	INVOICE TOTAL:	\$175.0		
10	2	3	XYZ INC.			
11	2	4	456 N. BROADWAY			
12	2	5	SALEM OR 98754			
13	2	6	-----			
14	3	1	ACCT: 34567	INVOIC		

Fig. 818

4) This step is critical. Copy the formulas in columns A and B and paste them back, using Paste Special Values. This will ensure that you can safely sort the data.

- 5) Sort the data by the Sequence number in column B. Your data will look like Fig. 819.

	A	B	C	D	E	F	G	H	I
1	Group	Seq	Text						
2	1	1	ACCT: 12345		INVOICE: 1010		DATE: 10/21/2005		
3	2	1	ACCT: 23456		INVOICE: 1011		DATE: 10/21/2005		
4	3	1	ACCT: 34567		INVOICE: 1012		DATE: 10/23/2005		
5	4	1	ACCT: 45678		INVOICE: 1013		DATE: 10/24/2005		
6	5	1	ACCT: 56789		INVOICE: 1014		DATE: 10/24/2005		
7	6	1	ACCT: 67890		INVOICE: 1015		DATE: 10/26/2005		
8	1	2	INVOICE TOTAL:		\$125.00				
9	2	2	INVOICE TOTAL:		\$175.00				
10	3	2	INVOICE TOTAL:		\$225.00				
11	4	2	INVOICE TOTAL:		\$425.00				
12	5	2	INVOICE TOTAL:		\$25.00				
13	6	2	INVOICE TOTAL:		\$185.00				
14	1	3	ABC CO						
15	2	3	XYZ INC.						
16	3	3	HANEY & ASSOCIATES						
17	4	3	FRIEDLINE FRYERS						
18	5	3	EASTERN POOLS						
19	6	3	BONNIE DOON						
20	1	4	123 S. MAIN STREET						
21	2	4	456 N. BROADWAY						

Fig. 819

You have now managed to intelligently segregate the data so that all similar records are together. A contiguous range of C2:C7 contains all of the first rows from each record. All of the line 1 records have three fields that really should be parsed into three separate columns. You can easily do this with the Text to Columns Wizard.

- 6) Select cells C2:C7. From the menu, select Data – Text to Columns. Select Fixed Width and Next, as shown in Fig. 820.

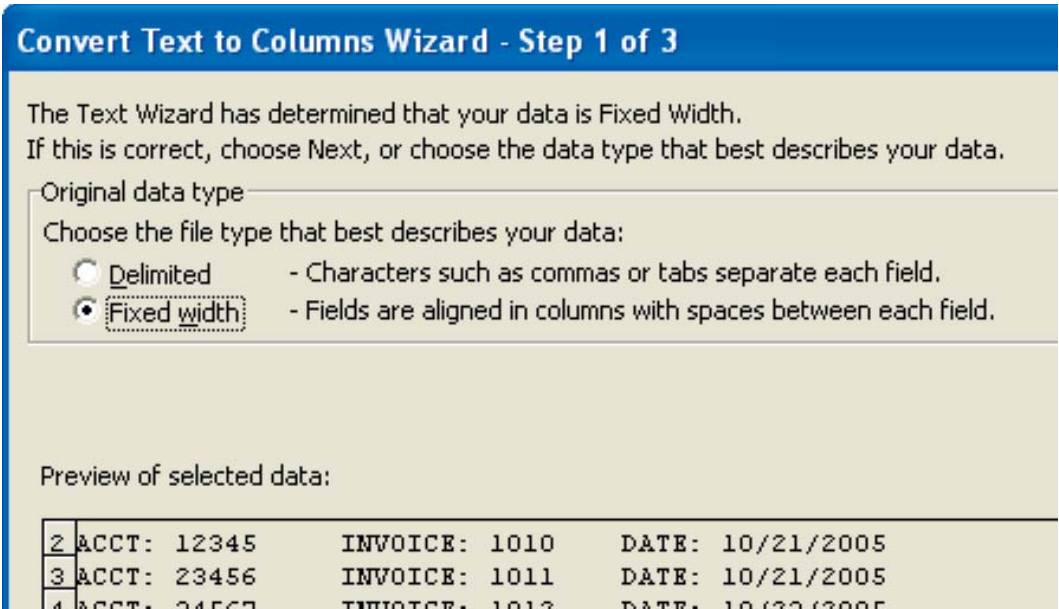


Fig. 820

- 7) Excel should properly guess where your columns are, as shown in Fig. 821. Click Next.

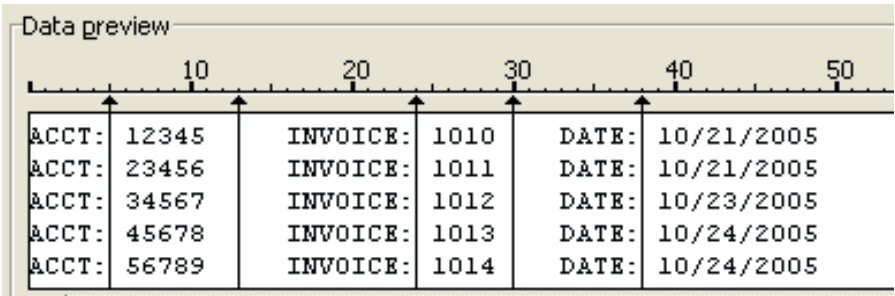


Fig. 821

- 8) Choose the heading for each column and define a data format. You don't really need the word ACCT each time, so choose to Skip the first, third, and fifth fields. Make the sixth field a date. When your information looks like Fig. 822, choose Finish.

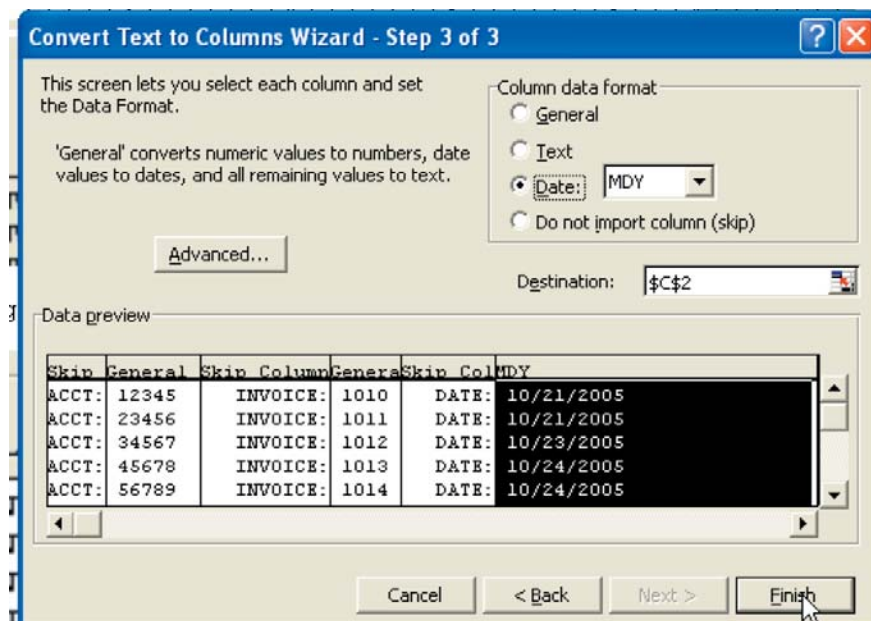


Fig. 822

- 9) You will have data in three columns of Group 1. As shown in Fig. 823, change the heading in C1 to be Acct, the heading in D1 to be Inv, and the heading in E1 to be Date.

	A	B	C	D	E	F
1	Group	Seq	Acct	Inv	Date	
2	1	1	12345	1010	10/21/2005	
3	2	1	23456	1011	10/21/2005	
4	3	1	34567	1012	10/23/2005	
5	4	1	45678	1013	10/24/2005	
6	5	1	56789	1014	10/24/2005	
7	6	1	67890	1015	10/26/2005	
8	1	2	INVOICE TOTAL:	\$125.00		
9	2	2	INVOICE TOTAL:	\$175.00		
10	3	2	INVOICE TOTAL:	\$225.00		
11	4	2	INVOICE TOTAL:	\$425.00		
12	5	2	INVOICE TOTAL:	\$25.00		
13	6	2	INVOICE TOTAL:	\$185.00		

Fig. 823

- 10) Next, select A8:C13. Cut and paste in F2. Add headings in F1:H1 of Group, Seq, and Total, as shown in Fig. 824.

E	F	G	H	I	J	K
Date	Group	Seq	Total			
10/21/2005	1	2	INVOICE	TOTAL:	\$125.00	
10/21/2005	2	2	INVOICE	TOTAL:	\$175.00	
10/23/2005	3	2	INVOICE	TOTAL:	\$225.00	
10/24/2005	4	2	INVOICE	TOTAL:	\$425.00	
10/24/2005	5	2	INVOICE	TOTAL:	\$25.00	
10/26/2005	6	2	INVOICE	TOTAL:	\$185.00	

Fig. 824

- 11) Select H2:H6 and choose Data – Text to Columns. In Step 1, select Fixed Width and choose Next. In Step 2, Excel offers to split your data into three fields. There is no need to have one column for the word Invoice and another column for the word Total. As shown in Fig. 825, double-click the line between Invoice and Total to delete the line.

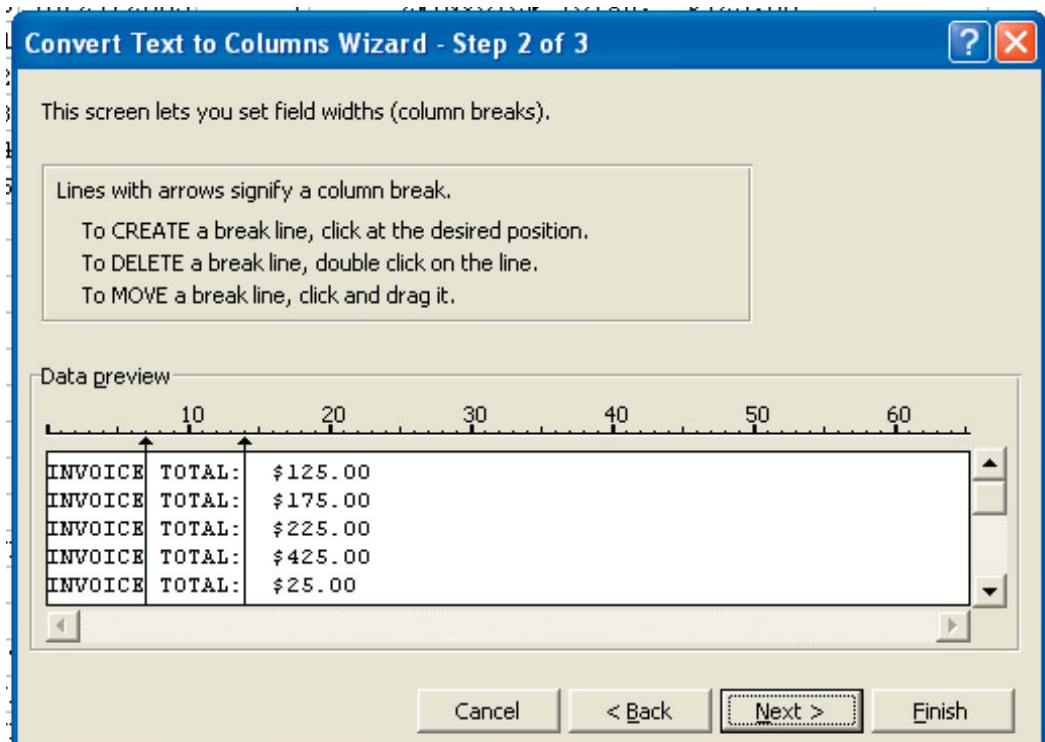
**Part
III**


Fig. 802

- 12) After double-clicking the first line, it is deleted. Choose Next, as shown in Fig. 826.

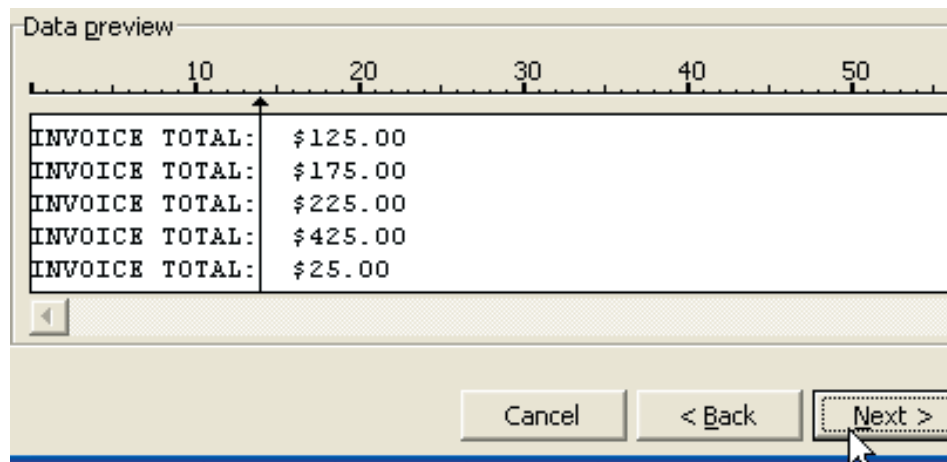


Fig. 826

- 13) In Step 3, choose to skip the field with “Invoice Total” and choose Finish, as shown in Fig. 827.

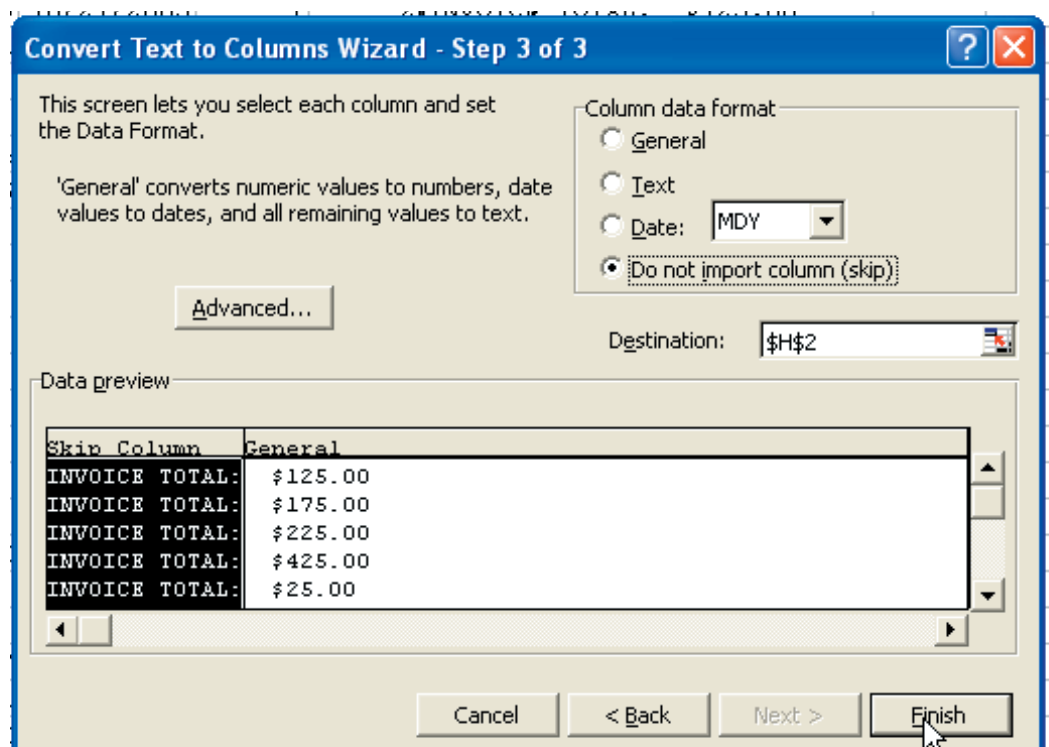


Fig. 827

- 14) Select the Group 3 records, as shown in Fig. 828.

13				
14	1	3	ABC CO	
15	2	3	XYZ INC.	
16	3	3	HANEY & ASSOCIATES	
17	4	3	FRIEDLINE FRYERS	
18	5	3	EASTERN POOLS	
19	6	3	BONNIE DOON	

Fig. 828

- 15) Copy them to I2. The headings in I1:K1 are Group, Seq, and Name, as shown in Fig. 829.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Group	Seq	Acct	Inv	Date	Group	Seq	Total	Group	Seq	Name		
2	1	1	12345	1010	10/21/2005	1	2	\$125.00	1	3	ABC CO		
3	2	1	23456	1011	10/21/2005	2	2	\$175.00	2	3	XYZ INC.		
4	3	1	34567	1012	10/23/2005	3	2	\$225.00	3	3	HANEY & ASSOCIATES		
5	4	1	45678	1013	10/24/2005	4	2	\$425.00	4	3	FRIEDLINE FRYERS		
6	5	1	56789	1014	10/24/2005	5	2	\$25.00	5	3	EASTERN POOLS		
7	6	1	67890	1015	10/26/2005	6	2	\$185.00	6	3	BONNIE DOON		
8													

Fig. 829

Part III

- 16) Select the Group 4 records. Cut and paste in L2.
- 17) Select the Group 5 records. Cut and paste in O2.
- 18) As shown in Fig. 830, the Group 6 records have no data – they are just dashed lines. You can delete these rows.

32	1	6	-----
33	2	6	-----
34	3	6	-----
35	4	6	-----
36	5	6	-----
37	6	6	-----
38			

Fig. 830

You now have all the fields, one line per record. You also have the words Group and Sequence taking up about five columns each. Before you delete the Group and Sequence columns, let's make sure that everything worked correctly. The Group numbers in columns A, F, I, L, and O should all match.

- 19) As shown in Fig. 831, in a blank column at the end, enter a large AND function as shown below. Copy this formula down to all rows:

=AND(A2=F2,F2=I2,I2=L2,L2=O2)

=AND(A2=F2,F2=I2,I2=L2,L2=O2)					
O	P	Q			R
Group	Seq	City	ST	ZIP	
1	5	SALEM	OH	44460	TRUE
2	5	SALEM	OR	98754	TRUE
3	5	SALEM	MA	12345	TRUE
4	5	SALEM	WV	32145	TRUE

Fig. 831

- 20) A value of True means that you have successfully put all of the Group 1 records back together. To test if all of the rows have True, enter the formula =AND(R2:R99) in cell R1. As shown in Fig. 832, if this formula is True, then you've crosschecked that all of the rows match up.

=AND(R2:R7)	
Q	R
ST ZIP	TRUE
OH 44460	TRUE
OR 98754	TRUE
MA 12345	TRUE
WV 32145	TRUE

Fig. 832

- 21) At this point, you can delete the columns that you don't need. As shown in Fig. 833, delete columns R, P, O, M, L, J, I, G, F, B, and A.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Group	Seq	Acct	Inv	Date	Group	Seq	Total	Group	Seq	Name	Group	Seq	Address	Group	Seq	City ST ZIP	
2	1	1	12345	1010	10/21/2005	1	2	\$125.00	1	3	ABC CO	1	4	123 S. MAIN STREET	1	5	SALEM OH 44460	TRUE
3	2	1	23456	1011	10/21/2005	2	2	\$175.00	2	3	XYZ INC.	2	4	456 N. BROADWAY	2	5	SALEM OR 98754	TRUE
4	3	1	34567	1012	10/23/2005	3	2	\$225.00	3	3	HANEY & ASSOCIATES	3	4	789 LUNDY LANE	3	5	SALEM MA 12345	TRUE
5	4	1	45678	1013	10/24/2005	4	2	\$425.00	4	3	FRIEDLINE FRYERS	4	4	987 KING CHURCH	4	5	SALEM WV 32145	TRUE

Fig. 833

Result: You now have a sortable, filterable, and reportable version of the original dataset. Each record consists of one row in Excel, as shown in Fig. 834.

	A	B	C	D	E	F	G
1	Acct	Inv	Date	Total	Name	Address	City ST ZIP
2	12345	1010	10/21/2005	\$125.00	ABC CO	123 S. MAIN STREET	SALEM OH 44460
3	23456	1011	10/21/2005	\$175.00	XYZ INC.	456 N. BROADWAY	SALEM OR 98754
4	34567	1012	10/23/2005	\$225.00	HANEY & ASSOCIATES	789 LUNDY LANE	SALEM MA 12345
5	45678	1013	10/24/2005	\$425.00	FRIEDLINE FRYERS	987 KING CHURCH	SALEM WV 32145
6	56789	1014	10/24/2005	\$25.00	EASTERN POOLS	654 FAIR AVE	SALEM IL 60187
7	67890	1015	10/26/2005	\$185.00	BONNIE DOON	321 PERSHING	SALEM IN 46875

Fig. 834

Summary: This process is convoluted. However, if you are presented with data as shown in the original example, the only way to quickly add up figures or to produce a report is to follow steps similar to the ones shown in this topic.

Commands Discussed: Data – Text to Columns

Functions Discussed: =IF(); =AND(); =LEFT()

ADD A CUSTOMER NUMBER TO EACH DETAIL RECORD

Problem: You've imported a dataset where they list the customer number once in column A and then have any number of invoice detail records. At the end of the first customer, they put the next customer number in column A and then have the detail records for that customer, as shown in Fig. 835. This is a common data format, but it is horrible in Excel. You cannot sort this dataset. You need to add the customer information to each record.

	A	B	C	D	E	F	G
1	Invoice	Date	Quantity	Product	Revenue	COGS	Profit
2	Acct A4651 Air Canada						
3	1533	31-Mar-04	300	ABC	5859.00	2541	3318
4	1559	19-Apr-04	200	XYZ	4948.00	2044	2904
5	1756	24-Aug-04	800	XYZ	17856.00	8176	9680
6	1828	7-Oct-04	100	DEF	2358.00	984	1374
7	Acct A8736 Ainsworth						
8	1404	3-Jan-04	500	XYZ	11240.00	5110	6130
9	1405	4-Jan-04	400	XYZ	9204.00	4088	5116
10	1414	10-Jan-04	900	XYZ	21465.00	9198	12267
11	1416	12-Jan-04	400	XYZ	9144.00	4088	5056

Fig. 835

Strategy:

- 1) Insert new columns A and B. Add headings of Acct and Customer, as shown in Fig. 836.

	A	B	C	D	E
1	Acct	Customer	Invoice	Date	Quantity
2			Acct A4651 Air Canada		
3			1533	31-Mar-04	300
4			1559	19-Apr-04	200
5			1756	24-Aug-04	800
6			1828	7-Oct-04	100
7			Acct A8736 Ainsworth		
8			1404	3-Jan-04	500

Fig. 836

Here is the basic logic of what Excel want to do in plain language: Look at the first four characters of column C. If they are equal to “Acct”, then this row has customer information. Take data from that cell and move it to column A. Otherwise, if the first four characters are anything other than “Acct”, use the same account information from the previous row’s column A.

The formula to do this for cell A2 is as follows:

=IF(LEFT(C2,4)="Acct",MID(C2,6,5),A1)

- 2) Enter this formula into cell A2 and copy it down through column A.

As shown in Fig. 837, as you copy this formula down, it does the job. In cell A2, the IF condition is true and data is extracted from C2. In cell A3, the condition is not true, so the value from A2 is used. Down in cell A7, a new customer number is found so the data from C7 is used in A7. Cells A8 through A59 get the customer number from A7.

**Part
III**

	A2		fx =IF(LEFT(C2,4)="Acct",MID(C2,6,5),A1)				
	A	B	C	D	E	F	
1	Acct	Customer	Invoice	Date	Quantity	Product	
2	A4651		Acct A4651 Air Canada				
3	A4651		1533	31-Mar-04	300	ABC	
4	A4651		1559	19-Apr-04	200	XYZ	
5	A4651		1756	24-Aug-04	800	XYZ	1
6	A4651		1828	7-Oct-04	100	DEF	
7	A8736		Acct A8736 Ainsworth				
8	A8736		1404	3-Jan-04	500	XYZ	1
9	A8736		1405	4-Jan-04	400	XYZ	
10	A8736		1414	10-Jan-04	900	XYZ	2

Fig. 837

Similar logic is needed in column B. In this case, though, you need to grab the customer name. You know that the word “Acct” and the space that follows it take up five characters. You know that your account number is another five characters and then there is a space before the customer name. Thus, you want to ignore the first 11 characters of cell C2.

You can use a formula of =MID(C2,12,50) to skip the first 11 characters and return the next 50 characters of the customer name. However, this formula will add spaces to the end of the customer name to ensure that you have 50 characters in the result. You really don't want all of those trailing spaces. The =TRIM() function will remove leading and trailing spaces from text.

Use =TRIM(MID(C2,12,50)) as the formula to extract a customer name. Use this formula as the True portion of the IF function. As shown in Fig. 838, the formula in B2, copied down through column B, is as follows:

=IF(LEFT(C2,4)="Acct",TRIM(MID(C2,12,50)),B1)

- 3) Enter this formula into cell B2 and copy it down through column B.

B65		fx =IF(LEFT(C65,4)="Acct",TRIM(MID(C65,12,50)),B64)					
	A	B	C	D	E	F	
1	Acct	Customer	Invoice	Date	Quantity	Product	Rev
62	B3529	Bell Canada	1662	24-Jun-04	200	DEF	40
63	B3529	Bell Canada	1773	31-Aug-04	800	XYZ	180
64	B3529	Bell Canada	1871	4-Nov-04	800	ABC	151
65	C434	Compton Petroleum	Acct C434	Compton Petroleum			
66	C4341	Compton Petroleum	1696	17-Jul-04	200	ABC	41
67	C4341	Compton Petroleum	1724	3-Aug-04	600	XYZ	139
68	C4341	Compton Petroleum	1782	8-Sep-04	100	DEF	20
69	C4341	Compton Petroleum	1851	21-Oct-04	500	DEF	112
70	C7849	Compaq	Acct C7849	Compaq			
71	C7849	Compaq	1642	6-Jun-04	400	XYZ	90
72	C7849	Compaq	1733	6-Aug-04	200	DEF	43
73	C7849	Compaq	1885	15-Nov-04	1000	ABC	172
74	C7849	Compaq	1900	24-Nov-04	400	XYZ	85
75	C8297	Chevron	Acct C8297	Chevron			
76	C8297	Chevron	1477	26-Feb-04	900	DEF	206
77	C8297	Chevron	1506	15-Mar-04	400	ABC	81
78	C8297	Chevron	1717	29-Jul-04	1000	ABC	182
79	C8297	Chevron	1835	12-Oct-04	300	XYZ	70
80	E3310	Exxon	Acct E3310	Exxon			
81	E3310	Exxon	1424	17-Jan-04	300	DEF	59

Fig. 838

You have now successfully filled in account and customer. You need to change these formulas to values.

- 4) Highlight columns A and B. Hit Ctrl+C to copy. Choose Edit – Paste Special – Values to convert the formulas to values.

The last task is to remove all of the customer heading rows. As you look for a method to isolate the heading rows, you will notice that heading rows are the only rows with blank cells in column D. You can move the blanks to the end of a dataset by sorting the data by column D.

- 5) Select the heading in D1. Hit the AZ sort button to sort ascending by Date. Any rows without a value in column D will automatically sort to the bottom of the dataset, as shown in Fig. 839.
- 6) From D1, press the End key and then the Down Arrow key twice. The cell pointer will be located on the first customer heading. Delete all of the rows below row 564.

	A	B	C	D	E	
1	Acct	Customer	Invoice	Date	Quantity	Proc
556	W7256	Wal-Mart	1955	24-Dec-04	200	XYZ
557	A8736	Ainsworth	1956	26-Dec-04	100	ABC
558	E3310	Exxon	1957	26-Dec-04	600	XYZ
559	S2571	SBC Communication	1958	26-Dec-04	500	XYZ
560	W7256	Wal-Mart	1959	26-Dec-04	700	DEF
561	S2328	Sun Life Financial	1960	27-Dec-04	500	ABC
562	G1394	General Motors	1961	28-Dec-04	600	ABC
563	G1394	General Motors	1962	28-Dec-04	600	ABC
564	G1394	General Motors	1963	28-Dec-04	900	ABC
565	A4651	Air Canada	Acct A4651	51 Air Canada		
566	A8736	Ainsworth	Acct A8736	Ainsworth		
567	B3529	Bell Canada	Acct B3529	Bell Canada		
568	C4341	Compton Petroleum	Acct C4341	Compton Petroleum		
569	C7849	Compaq	Acct C7849	Compaq		
570	C8297	Chevron	Acct C8297	Chevron		
571	E3310	Exxon	Acct E3310	Exxon		
572	F8417	Ford	Acct F8417	Ford		
573	G1394	General Motors	Acct G1394	General Motors		
574	G7759	Gildan Activewear	Acct G7759	Gildan Activewear		
575	H2221	HP	Acct H2221	HP		

Fig. 839

Result: You have a clean dataset with customer information on every row, as shown in Fig. 840. You can sort this data and otherwise use it for data analysis.

	A	B	C	D	E	F	G	H	I
1	Acct	Customer	Invoice	Date	Quantity	Product	Revenue	COGS	Profit
2	F8417	Ford	1401	1-Jan-04	1000	XYZ	22810.00	10220	12590
3	V8627	Verizon	1402	2-Jan-04	100	DEF	2257.00	984	1273
4	V8627	Verizon	1403	2-Jan-04	500	ABC	10245.00	4235	6010
5	A8736	Ainsworth	1404	3-Jan-04	500	XYZ	11240.00	5110	6130
6	A8736	Ainsworth	1405	4-Jan-04	400	XYZ	9204.00	4088	5116
7	G7759	Gildan Activewear	1406	4-Jan-04	800	DEF	18552.00	7872	10680
8	T5952	Texaco	1407	4-Jan-04	400	XYZ	9152.00	4088	5064

Fig. 840

Summary: A couple of formulas with IF functions help to snap this data into shape.

Commands Discussed: =IF(); =LEFT(); =MID()

USE A PIVOT TABLE TO SUMMARIZE DETAILED DATA

Problem: You have 50,000 rows of sales data, as shown in Fig. 841. You want to produce a summary report showing sales by region and product.

	A	B	C	D	E	F	G	H
1	Region	Product	Date	Customer	Quantity	Revenue	COGS	Profit
2	East	XYZ	1-Jan-04	Ford	1000	22810	10220	12590
3	Central	DEF	2-Jan-04	Verizon	100	2257	984	1273
4	East	ABC	2-Jan-04	Verizon	500	10245	4235	6010
5	Central	XYZ	3-Jan-04	Ainsworth	500	11240	5110	6130
6	Central	XYZ	4-Jan-04	Ainsworth	400	9204	4088	5116
7	East	DEF	4-Jan-04	Gildan Activewear	800	18552	7872	10680
8	East	XYZ	4-Jan-04	Texaco	400	9152	4088	5064

Fig. 841