SQL Step by Step:

An advanced tutorial for business users

Judy Loren
Independent Consultant

Gregory S. Barnes Nelson
ASG, Inc.

Abstract

Some excellent papers have been presented at SUGI and NESUG giving reasons to use SQL (Hermansen), overviews of SQL (Winn), selected features of SQL (Pass & Dickson, Peters), or behind the scene details of joins (Kent). This paper aims to solidify your understanding of the power of SQL by presenting an integrated series of examples that progress from the straightforward to the intricate. The paper will focus on coding but will pay some attention to performance issues and use of SQL via pass-thru to other databases.

Scope of this Paper

The purpose of this paper is to introduce advanced users of the SAS System to some of the more complicated tasks that can be accomplished using SQL. This is not intended to provide the reader with a blueprint of all the functionality found in SQL, but rather a focused exercise into the complexity of the language. Each of the examples is presented in order of common business questions that might have been asked.

The example we will be using throughout the paper will be data from the Northwinds Trading Company database that is distributed with Microsoft Access. The database is an order entry system for a company that sells a broad range of consumable/ food products. We selected this database for building our sample warehouse application for three reasons: (1) it is available to anyone who has MS-Access, (2) MS-Access is ODBC compliant and it was easy to get this data into SAS via SAS/ACCESS to ODBC, and (3) it provides a classic example of an operational system used in order processing that most of us have some familiarity with.

Data Structure

Just as a programmer would have to know what data elements are available in their operational data store, we too must understand the type of data that we have in order to answer questions about that data. Figure 1 shows out main subject entity-relationship diagram (ERD) that tells us basically what tables we have in our database.

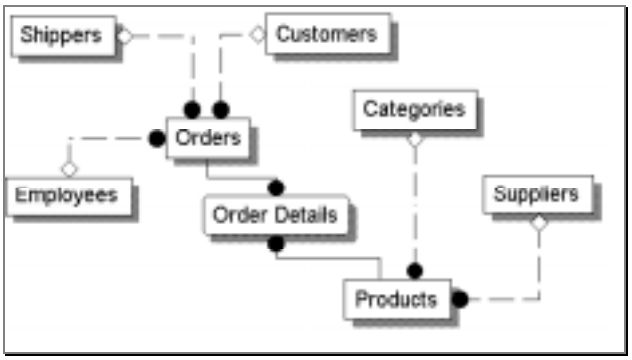

Figure 1. Main subject Area ERD

Note that this diagram shows the relationships between tables in our database. At the lowest level of detail we have information about specific orders that were placed at our fictitious company. For a detailed entity-relationship diagram, please refer to Appendix A.

From there, we can see the data in our system that support that level of detail, for example, we have information about our products, our customers, our employees in addition to having ancillary information about those things.

Now that we have at least a cursory understanding of our data, let's move forward and try to answer some basic business questions about this data. Let's turn our data into information.

From Data to Information

This paper is organized into 14 imcreasingly complex business questions. Each question will be presented along with a SQL solution. A brief description will follow each solution.

Q1 Manager asks, ingenuously, "What are sales for 1994?"

Concepts Introduced:

- PROC SQL
- Title statement
- Select (with as)
- Summary functions (down variables)
- Formating values
- From (Multiple datasets)
- Where (and)
- YEAR function

Solution

```
libname db "C:\WINDOWS\Nesug97\Sasdata";
proc sql;
title "Sales for 1994";
SELECT SUM(qty*unitprc) as sales format = dollar15.
FROM db.orddtls dtl,
     db.orders  ord
WHERE dtl.ordid = ord.ordid
  and year(orddate) = 1994;
```

Results

Sales for 1994

SALES
$339,885

Q2 Manager says, "How do I know whether this is good or bad? Show me how it's changed since last year!"

Concepts Introduced:

- CASE here is used to create a BY or CLASS variable-type analysis
- Summary functions (down and across variables)
- Calculated keyword

Solution

```
SELECT SUM(CASE WHEN year(orddate)=1994 then qty*unitprc
                ELSE 0
           END) as sales94 format dollar15.,
      SUM(CASE WHEN year(orddate)=1993 then qty*unitprc
                ELSE 0
           END) as sales93 format dollar15. ,
      CALCULATED sales94/CALCULATED sales93 -1 as pctchg
format=percent8.2
FROM db.orddtls dtl,
     db.orders ord
WHERE dtl.ordid=ord.ordid
  and year(orddate) between 1993 and 1994;
```

Results

Sales for 1994

SALES94	SALES93	PCTCHG
$339,885	$716,569	(52.57%)

Q3 Manager says, "Why are sales down so much from last year! Show me results by quarter!"

Concepts Introduced:

- Macro variables
- Macro programming

Solution

```
%macro calc(year,pct);
%do quarter = 1 %to 4;
  sum(case when year(orddate)=&year and
                month(orddate) between %eval(&quarter-1)*3+1 and
                                       %eval(&quarter-1)*3+3
           then (qty*unitprc)
      else 0
      end) as s&quarter&year format= dollar15.,
%end;
%do quarter=1 %to 4;
%if "&pct"="y" %then %do;
    %let lastyear=%eval(&year-1);
    (calculated s&quarter&year/calculated s&quarter&lastyear) -1
      as p&quarter&year format=percent15.2
%end;
%if "&year" = "1994" and "&quarter" ne "4" %then ,  ;
%end;

%mend;
proc sql;
title "Sales Growth 1993 to 1994 by Quarter";
SELECT %calc(1993,n)
       %calc(1994,y)
FROM db.orddtls dtl,
     db.orders ord
WHERE dtl.ordid=ord.ordid;
```

Results

Sales Growth 1993 to 1994 by Quarter

```
S11993       S21993       S31993          S41993
S11994       S21994       S31994          S41994
P11994       P21994       P31994          P41994
------------------------------------------------
$142,983     $145,644     $183,036        $244,905
$339,885           $0           $0              $0
137.71%    (  100.00%)  ( 100.00%)  (     100.00%)
```

Q4 **Manager says, "Oh, that's right, we're only 1 quarter into 1994. Silly me. Show me the latest year vs the previous 12 months. And let's look at whether we've been able to add customers to our base. And check whether we've lost any customers. In fact, show me whether customers are increasing their sales or decreasing their sales."**

Concepts Introduced:

- Coalesce
- Full Join
- Group by
- Order
- Subquery
- CASE logic to create a new variable (flag)

Solution

```
TITLE "Customer Growth Analysis";
SELECT COALESCE(list93.custid, list94.custid) as custid,
       case when sales94 > sales93 and sales93 > 0 then "Growing"
            when sales94 > sales93 and sales93 <=0 then "New"
            when sales94 < sales93 and sales94 > 0 then
"Shrinking"
            when sales94 <= 0 then "Lost"
       else "Not Known"
        end as type label="Customer Growth Category"
FROM (SELECT custid, SUM(qty*unitprc) as sales93
      FROM   db.orders ord, db.orddtls dtl
      WHERE  ord.ordid=dtl.ordid
        and  ((year(orddate)=1993 and  qtr(orddate) =1)
           or (year(orddate)=1992 and qtr(orddate) gt 1))
      GROUP BY custid) as list93
      FULL JOIN
     (SELECT custid, SUM(qty*unitprc) as sales94
      FROM   db.orders ord, db.orddtls dtl
      WHERE  ord.ordid=dtl.ordid
        and  ((year(orddate)=1994 and  qtr(orddate) =1)
           or (year(orddate)=1993 and qtr(orddate) gt 1))
      GROUP BY custid) as list94
    ON list93.custid = list94.custid
ORDER BY type, custid
;
```

Results

```
                Customer Growth Analysis

                                   Customer
                                   Growth
                         CUSTID    Category
                         ----------------
                         ANATR    Growing
                         ANTON    Growing
                         AROUT    Growing
                                     .
                                     .
                                     .
                         WHITC    Growing
                         WOLZA    Growing
                         CENTC    Lost
                         ALFKI    New
                         ARANR    New
                         LACOR    New
                         LETSS    New
                         THECR    New
                         TRAIH    New
                         WILMK    New
                         BLONP    Shrinking
                         COMMI    Shrinking
                         CONSH    Shrinking
                                     .
                                     .
                                     .
                         WANDK    Shrinking
                         WARTH    Shrinking
```

Q5 **Manager says, "What are these ids? I need the company name."**

Concepts Introduced:

- Multi-table outer join (full join ON var=var, table)

Solution

```
SELECT COALESCE(list93.custid, list94.custid) as custid,
       company label='Company Name',
       case when sales94 > sales93 and sales93 > 0 then "Growing"
            when sales94 > sales93 and sales93 <=0 then "New"
            when sales94 < sales93 and sales94 > 0 then
"Shrinking"
            when sales94 <= 0 then "Lost"
       else "Not Known"
        end as type label="Customer Growth Category"
FROM (SELECT custid, SUM(qty*unitprc) as sales93
      FROM   db.orders ord, db.orddtls dtl
      WHERE  ord.ordid=dtl.ordid
        and  ((year(orddate)=1993 and  qtr(orddate) =1)
           or (year(orddate)=1992 and qtr(orddate) gt 1))
      GROUP BY custid) as list93
      FULL JOIN
     (SELECT custid, SUM(qty*unitprc) as sales94
      FROM   db.orders ord, db.orddtls dtl
      WHERE  ord.ordid=dtl.ordid
        and  ((year(orddate)=1994 and  qtr(orddate) =1)
           or (year(orddate)=1993 and qtr(orddate) gt 1))
      GROUP BY custid) as list94
    ON list93.custid = list94.custid,
    db.custs cust
WHERE CALCULATED custid = cust.custid
ORDER BY type, custid
;
```

Results

```
Customer Growth Analysis

                                                    Customer
                                                    Growth
      CUSTID  Company Name                          Category
      ----------------------------------------------------------
      ANATR   Ana Trujillo Emparedados y helados    Growing
      ANTON   Antonio Moreno Taquería               Growing
      AROUT   Around the Horn                       Growing
```

```
                        .
                        .
                        .
WHITC   White Clover Markets                    Growing
WOLZA   Wolski  Zajazd                         Growing
CENTC   Centro comercial Moctezuma              Lost
ALFKI   Alfreds Futterkiste                     New
FRANR   France restauration                     New
LACOR   La corne d'abondance                    New
LETSS   Let's Stop N Shop                       New
THECR   The Cracker Box                         New
TRAIH   Trail's Head Gourmet Provisioners       New
WILMK   Wilman Kala                             New
BLONP   Blondel père et fils                    Shrinking
COMMI   Comércio Mineiro                        Shrinking
CONSH   Consolidated Holdings                   Shrinking
                        .
                        .
                        .

WANDK   Die Wandernde Kuh                       Shrinking
WARTH   Wartian Herkku                          Shrinking
```

Q6 Manager says, "I can't tell anything from this list. Show me where the companies are located."

Concepts Introduced:

- Reset flow

Solution

```
RESET FLOW;
SELECT COALESCE(list93.custid, list94.custid) as custid,
       company label='Company Name',
       country label='Company Location',
       case when sales94 > sales93 and sales93 > 0 then "Growing"
            when sales94 > sales93 and sales93 <=0 then "New"
            when sales94 < sales93 and sales94 > 0 then
"Shrinking"
            when sales94 <= 0 then "Lost"
       else "Not Known"
        end as type label="Customer Growth Category"
FROM (SELECT custid, SUM(qty*unitprc) as sales93
      FROM   db.orders ord, db.orddtls dtl
      WHERE  ord.ordid=dtl.ordid
        and  ((year(orddate)=1993 and  qtr(orddate) =1)
           or (year(orddate)=1992 and qtr(orddate) gt 1))
      GROUP BY custid) as list93
      FULL JOIN
     (SELECT custid, SUM(qty*unitprc) as sales94
      FROM   db.orders ord, db.orddtls dtl
      WHERE  ord.ordid=dtl.ordid
        and  ((year(orddate)=1994 and  qtr(orddate) =1)
           or (year(orddate)=1993 and qtr(orddate) gt 1))
      GROUP BY custid) as list94
    ON list93.custid = list94.custid,
    db.custs cust
WHERE CALCULATED custid = cust.custid
ORDER BY type, custid
;
```

Results

```
Customer Growth Analysis

Customer
                                    Company          Growth
        CUSTID  Company Name          Location         Category
        ------------------------------------------------------------
        ANATR   Ana Trujillo          Mexico           Growing
                Emparedados
                y helados
        ANTON   Antonio Moreno        Mexico           Growing
                Taquería
        AROUT   Around the Horn       UK               Growing
                        .
                        .
```

```
                        .
                        .
WHITC   White Clover Markets    USA             Growing
WOLZA   Wolski  Zajazd          Poland          Growing
CENTC   Centro comercial        Mexico          Lost
        Moctezuma
ALFKI   Alfreds Futterkiste     Germany         New
FRANR   France restauration     France          New
LACOR   La corne d'abondance    France          New
LETSS   Let's Stop N Shop       USA             New
THECR   The Cracker Box         USA             New
TRAIH   Trail's Head Gourmet    USA             New
        Provisioners
WILMK   Wilman Kala             Finland         New
BLONP   Blondel père et fils    France          Shrinking
COMMI   Comércio Mineiro        Brazil          Shrinking
CONSH   Consolidated            UK              Shrinking
        Holdings
                        .
                        .
                        .

WANDK   Die Wandernde Kuh       Germany         Shrinking
WARTH   Wartian Herkku          Finland         Shrinking
```

Q7 Manager says, "Well this list is too long to really tell what's going on. Just show me how many are in each growth category by country."

Concepts Introduced:

- Create view
- Subqueries (instead of view?)

Solution

```
RESET FLOW ;
create view growth as
SELECT COALESCE(list93.custid, list94.custid) as custid,
       company label='Company Name',
       country label='Company Location',
       case when sales94 > sales93 and sales93 > 0 then "Growing"
            when sales94 > sales93 and sales93 <=0 then "New"
            when sales94 < sales93 and sales94 > 0 then
"Shrinking"
            when sales94 <= 0 then "Lost"
       else "Not Known"
        end as type label="Customer Growth Category"
FROM (SELECT custid, SUM(qty*unitprc) as sales93
      FROM   db.orders ord, db.orddtls dtl
      WHERE  ord.ordid=dtl.ordid
        and  ((year(orddate)=1993 and  qtr(orddate) =1)
           or (year(orddate)=1992 and qtr(orddate) gt 1))
      GROUP BY custid) as list93
      FULL JOIN
     (SELECT custid, SUM(qty*unitprc) as sales94
      FROM   db.orders ord, db.orddtls dtl
      WHERE  ord.ordid=dtl.ordid
        and  ((year(orddate)=1994 and  qtr(orddate) =1)
           or (year(orddate)=1993 and qtr(orddate) gt 1))
      GROUP BY custid) as list94
    ON list93.custid = list94.custid,
    db.custs cust
WHERE CALCULATED custid = cust.custid
ORDER BY type, country
;

select count(*) as cos label='Companies', country, type
from growth
group by country, type
order by country
;
```

Results

```
Customer Growth Analysis
```

Companies	Company Location	Customer Growth Category
3	Argentina	Growing
1	Austria	Growing
1	Austria	Shrinking
1	Belgium	Growing
1	Belgium	Shrinking
7	Brazil	Growing
2	Brazil	Shrinking
1	Canada	Growing
2	Canada	Shrinking
1	Denmark	Growing
1	Denmark	Shrinking
1	Finland	New
1	Finland	Shrinking
5	France	Growing
2	France	New
3	France	Shrinking
8	Germany	Growing
1	Germany	New
2	Germany	Shrinking
1	Ireland	Growing
3	Italy	Growing
2	Mexico	Growing
1	Mexico	Lost
2	Mexico	Shrinking
1	Norway	Growing
1	Poland	Growing
1	Portugal	Growing
1	Portugal	Shrinking
2	Spain	Growing
2	Spain	Shrinking
2	Sweden	Growing
2	Switzerland	Growing
4	UK	Growing
3	UK	Shrinking
8	USA	Growing
3	USA	New
2	USA	Shrinking
2	Venezuela	Growing

Q8

Manager says, "I can't take any action on this! I want to know which countries have increased sales and which have lost sales in the last 12 months! And show it to me in some kind of order so I know which countries are the best and which are the worst!"

Concepts Introduced:

- summary statistics (across variables)
- DESC on GROUP BY

Solution

```
TITLE "Geographic Growth Analysis";
SELECT COALESCE(list93.country, list94.country) as country
              label='Company Location',
       sales93 format=dollar15. label='1 Year Previous',
       sales94 format=dollar15. label='Most Recent 4 Quarters',
       sales94/sales93 -1 as pctchg label='Percent Change'
format=percent15.2

FROM (SELECT country, SUM(qty*unitprc) as sales93
      FROM   db.orders ord, db.orddtls dtl, db.custs  cust
      WHERE  ord.ordid=dtl.ordid
        and  ((year(orddate)=1993 and  qtr(orddate) =1)
           or (year(orddate)=1992 and qtr(orddate) gt 1))
        and  ord.custid = cust.custid
      GROUP BY country) as list93
      FULL JOIN
     (SELECT country, SUM(qty*unitprc) as sales94
      FROM   db.orders ord, db.orddtls dtl, db.custs cust
      WHERE  ord.ordid=dtl.ordid
        and  ((year(orddate)=1994 and  qtr(orddate) =1)
           or (year(orddate)=1993 and qtr(orddate) gt 1))
        and  cust.custid = ord.custid
      GROUP BY country) as list94
    ON list93.country = list94.country

ORDER BY pctchg desc
;
```

Results

Geographic Growth Analysis

Company Location	1 Year Previous	Most Recent 4 Quarters	Percent Change
Argentina	$988	$7,131	621.69%
Poland	$459	$3,073	569.49%
Norway	$1,258	$4,477	255.75%
Spain	$4,645	$15,993	244.34%
Sweden	$17,051	$47,115	176.32%
Switzerland	$9,173	$24,066	162.36%
Italy	$5,371	$13,011	142.25%
Germany	$75,215	$174,479	131.97%
Brazil	$36,685	$79,421	116.50%
Ireland	$19,574	$40,344	106.11%
USA	$93,541	$188,567	101.59%
Austria	$50,436	$89,061	76.58%
Venezuela	$22,712	$39,567	74.21%
UK	$23,848	$40,595	70.23%
Finland	$8,724	$11,055	26.72%
France	$38,776	$48,815	25.89%
Belgium	$17,471	$21,375	22.35%
Denmark	$17,014	$17,768	4.43%
Portugal	$6,159	$6,310	2.46%
Mexico	$13,137	$12,289	(6.46%)
Canada	$32,125	$28,958	(9.86%)

Q9

Manager says, "Well how can I tell what's going on if I don't know what's selling? Maybe the countries doing more business like some of our product categories better than others. Show me the growth in business by product category, and index the percent change to the total company growth rate so I can see which ones are driving our growth!

Concepts Introduced:

- Reset noprint
- How to fill a macro variable from SQL
- Putting grand total on subtotal lines

Solution

```
libname db "C:\WINDOWS\Nesug97\Sasdata";
proc sql;
reset flow;
reset noprint;
TITLE "Growth in Product Categories Compared to Overall Growth";

SELECT SUM(CASE WHEN year(orddate)=1994
                  or (year(orddate)=1993 and qtr(orddate) gt 1)
                then qty*unitprc
                  ELSE 0
                END) into :tot94
    FROM db.orddtls dtl,
         db.orders ord
    WHERE dtl.ordid=ord.ordid
     and year(orddate) between 1993 and 1994 ;
  SELECT
      SUM(CASE WHEN  (year(orddate)=1993 and qtr(orddate)=1) or
                     (year(orddate)=1992 and qtr(orddate) gt 1)
```

```
              then dtl.qty*dtl.unitprc
               ELSE 0
          END) into :tot93
     FROM db.orddtls dtl,
          db.orders ord
     WHERE dtl.ordid=ord.ordid
      and year(orddate) between 1992 and 1993;
reset print;
SELECT cat.catname format=$15. label='Product Category',
      SUM(CASE WHEN year(orddate)=1994
                    or (year(orddate)=1993 and qtr(orddate) gt 1)
               then dtl.qty*dtl.unitprc
               ELSE 0
          END) as sales94 format dollar10. label='Sales Q2 93-Q1
94',
      SUM(CASE WHEN (year(orddate)=1993 and qtr(orddate)=1) or
                    (year(orddate)=1992 and qtr(orddate) gt 1)
               then dtl.qty*dtl.unitprc
               ELSE 0
          END) as sales93 format dollar10. label='Sales Q2 92-Q1
93' ,
      CALCULATED sales94/CALCULATED sales93 -1 as catchg
format=percent8.1
                label='Category Percent Change',
      &tot94/&tot93 -1 as pctchg format=percent8.1
                label='Company Percent Change',
      CALCULATED catchg/CALCULATED pctchg * 100 as index
format=8.1
                label='Category Change Index to Company'


FROM db.orddtls dtl,
     db.orders ord,
     db.cats cat,
     db.products prod

WHERE dtl.ordid=ord.ordid
  and dtl.prodid=prod.prodid
  and prod.catid=cat.catid
  and year(orddate) between 1992 and 1994
GROUP BY cat.catname
;
```

Results

Growth in Product Categories Compared to Overall Growth

Product Category	Sales Q2 93-Q1 94	Sales Q2 92-Q1 93	Category Percent Change	Company Percent Change	Category Change Index to Company
Beverages	$184,024	$113,172	62.6%	84.8%	73.8
Condiments	$73,010	$44,832	62.9%	84.8%	74.1
Confections	$112,087	$70,615	58.7%	84.8%	69.3
Dairy Products	$173,979	$86,711	100.6%	84.8%	118.7
Grains/Cereals	$70,250	$34,110	106.0%	84.8%	125.0
Meat/Poultry	$120,137	$72,652	65.4%	84.8%	77.1
Produce	$74,727	$31,212	139.4%	84.8%	164.4
Seafood	$105,255	$41,056	156.4%	84.8%	184.4

Q11

Manager says, "We have to market our product categories more aggressively. Let's hire a consultant to do some real analysis. You provide the consultant with a table showing which products are ordered together. Then we can come up with some upselling guidelines."

Concepts Introduced:

- Reset outobs=50
- create table
- reflexive join (table with itself)

Solution

```
libname db "C:\WINDOWS\Nesug97\Sasdata";
proc sql;
CREATE table db.pairs as
SELECT a.prodid as prod1,
       b.prodid as prod2,
       c.custid
FROM   db.orddtls a,
       db.orddtls b,
       db.orders c
WHERE  a.ordid=b.ordid
  and  a.ordid=c.ordid
  and  a.prodid < b.prodid
ORDER BY custid
;

reset outobs=50;
select * from db.pairs;
run;
```

Results

OBS	PROD1	PROD2	CUSTID
1	51	55	ALFKI
2	5	51	ALFKI
3	5	55	ALFKI
4	28	39	ALFKI
5	28	46	ALFKI
6	59	77	ALFKI
7	39	46	ALFKI
8	3	76	ALFKI
9	58	71	ALFKI
10	6	28	ALFKI
11	11	13	ANATR
12	11	19	ANATR
13	11	72	ANATR
14	13	72	ANATR
15	19	72	ANATR
16	13	19	ANATR
17	69	70	ANATR
18	14	42	ANATR
19	42	60	ANATR
20	14	60	ANATR
21	11	40	ANTON
22	17	53	ANTON
23	11	59	ANTON
24	43	48	ANTON
25	57	59	ANTON
26	40	59	ANTON
27	17	34	ANTON
28	34	53	ANTON
29	40	57	ANTON
30	11	57	ANTON
31	33	66	ANTON

Q10

Manager says, "I can't wait for that consultant to finish his analysis. Just show me the things that are ordered together most often.

Concepts Introduced:

- summary functions
- embedded use of summary statistics to select records
- subquery to put grand total stats on subtotal records

Solution

```
libname db "C:\WINDOWS\Nesug97\Sasdata";
```

```
proc sql;

create table combos as
SELECT count(*) as occurs, prod1, prod2
FROM   db.pairs
GROUP BY prod1,prod2;

SELECT b.proddesc as proda label='Product 1'
     , c.proddesc as prodc label='Product 2'
     , occurs label='Times Ordered Together'
FROM  combos a,
      db.products b,
      db.products c
WHERE a.prod1=b.prodid
  and a.prod2=c.prodid
  and occurs >
(SELECT mean(occurs) + std(occurs) *3
 FROM combos)
ORDER BY occurs desc
;
```

Results

```
Times
           Product 1                                 Ordered
           Product 2                                Together
           -------------------------------------------------
           Sir Rodney's Scones
           Maple Syrup                                     8

           Pavlova Meringue Dessert
           Gorgonzola Telino                               7

           Pavlova Meringue Dessert
           Pierrot Camembert                               7

           Pavlova Meringue Dessert
           Rhönbräu Beer                                   7

           Sir Rodney's Scones
           Manjimup Dried Apples                           6

           Pavlova Meringue Dessert
           Sugar Pie                                       6

           Gorgonzola Telino
           Giovanni's Mozzarella                           6

           Pierrot Camembert
           Fløtemys Cream Cheese                           6

           Thüringer Sausage
           Courdavault Raclette Cheese                     6

           Tibetan Barley Beer
           Pavlova Meringue Dessert                        6

           Nord-Ost White Herring
           Pork Pie                                        6
```

Q13

Manager says, "I went to a Marketing conference and got this great idea. Let's categorize Customers based on the products they order. Customers who buy mostly Beverages, Condiments, and Confections are junk food addicts. But we'll call them miscellaneous. Customers who order mostly Dairy Products and Produce are farmers at heart. Customers who like Grains/Cereals and Produce are vegan. And customers who order mostly Meat/Poultry and Seafood are protein heads. How many customers do we have who fit these definitions?"

Concepts Introduced:

- between
- in
- putting subtotals on detail records

Solution

```
libname db "C:\WINDOWS\Nesug97\Sasdata";
proc sql;
title "Customer Groups by Preferred Category of Product";
CREATE VIEW custgrps as
SELECT gtot.custid,
       custtot,
       sum(case when catid between 1 and 3 then qty*dtl.unitprc
             else 0 end) as misc,
       sum(case when catid in (4,7) then qty*dtl.unitprc
             else 0 end) as farm,
       sum(case when catid in (5,7) then qty*dtl.unitprc
             else 0 end) as vegan,
       sum(case when catid in (6,8) then qty*dtl.unitprc
             else 0 end) as protein
FROM  (SELECT custid,
              sum(qty*unitprc) as custtot
       FROM   db.orders ord,
              db.orddtls dtl
       WHERE ord.ordid=dtl.ordid
       GROUP BY custid) as gtot,
       db.orders ord,
       db.orddtls dtl,
       db.products prod
WHERE  gtot.custid=ord.custid
   and ord.ordid=dtl.ordid
   and dtl.prodid=prod.prodid
GROUP BY gtot.custid;

SELECT count(*) as compns,
       custgrp
FROM  (SELECT custid,
              case when misc/custtot > .5 then "Misc   "
                    when farm/custtot > .5 then "Farm   "
                    when vegan/custtot > .5 then "Vegan  "
                    when protein/custtot > .5 then "Protein"
                    else "Omni"
                    end as custgrp
       FROM custgrps)
GROUP BY custgrp;
```

Results

```
Customer Groups by Preferred Category of Product

                                 COMPNS  CUSTGRP
                               -----------------
                                    4  Farm
                                   30  Misc
                                   47  Omni
                                    7  Protein
                                    2  Vegan
```

Q12

Manager says "I'll look at this later [since it obviously didn't work]. But right now we've got real problems. I just found out that one of our confounded warehouses had a problem with a freezer case in January. We've got to find all the customers who got any pies since the first of the year so we can recall the

product, send them an apology and all that. First tell me what kinds of pies we sell."

Concepts Introduced:

- like %
- UPCASE

Solution 1

```
select prod.proddesc
from db.products prod
where upcase(prod.proddesc) like "%PIE%"
;
```

Results 1

```
PRODDESC
----------------------------
Perth Meat Pies
Pork Pie
Shepard's Pie
Pierrot Camembert
Sugar Pie
```

Note: The Pierrot Camembert shouldn't be selected. Have to use the leading space on "% PIE%"

Solution 2

```
select prod.proddesc
from db.products prod
where upcase(prod.proddesc) like "% PIE%"
;
```

Results 2

```
PRODDESC
---------------
Perth Meat Pies
Pork Pie
Shepard's Pie
Sugar Pie
```

Q14

Manager say, "Now give me a nice list of all the customers who got any of these pies since January 1. Make sure you give me addresses, phone numbers, contact names, the works."

Concepts Introduced:

- PROC FORMS. SAS programmers shouldn't forget their roots in other PROCs

Solution

```
proc sql;
title 'Customers Who Received Pies after Jan. 1, 1994';
create view piecusts as
select distinct cust.*
from db.orders ord,
     db.orddtls dtl,
     db.products prod,
     db.custs cust
where ord.ordid=dtl.ordid
  and dtl.prodid=prod.prodid
  and ord.custid=cust.custid
  and upcase(prod.proddesc) like "% PIE%"
  and year(ord.shipdate) = 1994
;
quit;

proc forms data=piecusts lines=6 nacross=2 width=40;
line 1 company;
line 2 contact;
line 3 conttitl;
line 4 address;
line 5 city region zip country / pack;
line 6 phone;

run;
```

Results

```
Customers Who Received Pies after Jan. 1, 1994

Bon app'                        Bottom-Dollar Markets
Laurence Lebihan                Elizabeth Lincoln
Owner                           Accounting Manager
12, rue des Bouchers            23 Tsawassen Blvd.
Marseille 13008 France          Tsawassen BC T2F 8M4 Canada
91.24.45.40                     (604) 555-4729

Chop-suey Chinese               Eastern Connection
Yang Wang                       Ann Devon
Owner                           Sales Agent
Hauptstr. 29                    35 King George
Bern 3012 Switzerland           London WX3 6FW UK
0452-076545                     (71) 555-0297

Ernst Handel                    Folk och fä HB
Roland Mendel                   Maria Larsson
Sales Manager                   Owner
Kirchgasse 6                    Åkergatan 24
Graz 8010 Austria               Bräcke S-844 67 Sweden
7675-3425                       0695-34 67 21

Godos Cocina Típica             Island Trading
José Pedro Freyre               Helen Bennett
Sales Manager                   Marketing Manager
C/ Romero, 33                   Garden House Crowther Way
Sevilla 41101 Spain             Hedge End Lancashire LA9 PX8 UK
(95) 555 82 82                  (24) 555-8888

Königlich Essen                 Lonesome Pine Restaurant
Philip Cramer                   Fran Wilson
Sales Associate                 Sales Manager
                  Maubelstr. 90                     89
                  Chiaroscuro Rd.
                  Brandenburg 14776 Germany         Portland
                  OR 97219 USA
                  0555-09876                        (503)
                  555-9573

                  QUICK-Stop
                  Rattlesnake Canyon Grocery
Horst Kloss                     Paula Wilson
Accounting Manager              Assistant Sales Rep
Taucherstraße 10                2817 Milton Dr.
Cunewalde 01307 Germany         Albuquerque NM 87110 USA
0372-035188                     (505) 555-5939

Richter Supermarkt              Save-a-lot Markets
Michael Holz                    Jose Pavarotti
```

```
Sales Manager                    Sales Representative
Grenzacherweg 237                187 Suffolk Ln.
Genève 1203 Switzerland          Boise ID 83720 USA
0897-034214                      (208) 555-8097

Suprêmes délices                 Tortuga Restaurante
Pascale Cartrain                 Miguel Angel Paolino
Accounting Manager               Owner
Boulevard Tirou, 255             Avda. Azteca 123
Charleroi B-6000 Belgium         México D.F. 05033 Mexico
(071) 23 67 22 20                (5) 555-2933

White Clover Markets             Wolski  Zajazd
Karl Jablonski                   Zbyszek Piestrzeniewicz
Owner                            Owner
305 - 14th Ave. S. Suite 3B      ul. Filtrowa 68
Seattle WA 98128 USA             Warszawa 01-012 Poland
(206) 555-4112                   (26) 642-7012
```

Conclusion

The goal of this paper was to provide the reader with a practical guide to using SQL for answering some common business questions. The advanced nature of this paper allowed us to show how SQL can be used to perform some fairly complex analysis with fairly straightforward code. Our hope is that when you read this paper, it will provoke some thought on your part on how you might use SQL for more than just joins and simple queries.

Acknowledgments

We would like to give special thanks to Paul Kent and Sigurd Hermansen for their assistance in reviewing this manuscript.

Bibliography

"Fun with the SQL Procedure--An Advanced Tutorial", Lewis Church and Paul Kent, NESUG '93 p. 29

"Ten Good Reasons to Learn SAS Software's SQL Procedure", Sig Hermansen, SUGI 22, p. 180

"Introduction to Using PROC SQL" Tom Winn SUGI 22 p. 383

"SELECT ITEMS FROM PROC.SQL WHERE ITEMS > BASICS" Alan Dickson and Ray Pass SUGI 21 p. 227

"An SQL Tutorial: Some Random Tips" Paul Kent SUGI 21 p. 237

Address

Questions pertaining to this article should be addressed to:

Judith F. Loren
P.O. Box 306
Cumberland Center, Maine 04021
(207) 829-6246

Gregory S. Barnes Nelson
2000 Regency Parkway, Suite 355
Cary, NC 27511
919-467-0505
919-467-2469 (Fax)
gbn@asg-inc.com

Appendix

Entity-Relationship Diagram for the Northwinds Trading Company Database.