Sample Work with SQL in SAS

Heidi Kang

11/02/2018

Problem 1a

I used proc sql to start the process. I used a select clause and selected Employee_Name, Employee_Gender, department, Job_Title, and Salary since that is the information that was requested in the problem. I used a where clause so that the observations would only include females that were in a management department with a salary between $37,951 and $91,900.

```
libname stat440'/home/hmkang30/my_courses/kinson2/stat440_data';
 proc sql ;
 select Employee_Name, Employee_Gender, Department, Job_Title,
 Salary
 from stat440.employee_roster
 where Employee_Gender in ('F') and Department like
 '% Management' and Salary >= 37951 and Salary <= 91900;
quit;
```

There are a lot of females in the Group HR management department. Most of the employees make around $40,000 or $50,000 but there are a few who make over $80,000.

Employee Name	Employee Gender	Department	Job Title	Annual Salary
Kenisha Winge	F	Concession Management	Concession Manager	$43,635
Salley Amos	F	Logistics Management	Logistics Coordinator II	$42,570
Cynthia Racine	F	Logistics Management	Senior Logistics Manager	$85,495
Anglar Kornblith	F	Logistics Management	Pricing Manager	$56,385
Jill Leacock	F	Logistics Management	Senior Logistics Manager	$80,070
Nikeisha Kokoszka	F	Accounts Management	Financial Controller III	$51,950
Janelle Kempster	F	Accounts Management	Auditing Manager	$53,400

Employee Name	Employee Gender	Department	Job Title	Annual Salary
Julia Pascoe	F	Accounts Management	Auditor I	$43,930
Sue El-Amin	F	Group HR Management	HR Specialist II	$45,155
Tanya Thompson	F	Group HR Management	HR Analyst II	$41,580
Angela Gardner	F	Group HR Management	HR Specialist I	$43,650
Jennifer Eggleston	F	Group HR Management	HR Analyst II	$43,690
Kimberly Walcott	F	Group HR Management	HR Manager	$62,995
Rilma Sines	F	Group HR Management	Recruitment Manager	$63,915
Damesha Donnell	F	Group HR Management	Training Manager	$48,335
Renee Capachietti	F	Sales Management	Sales Manager	$83,505

Problem 1b

I used a select clause to select Employee_Name, Job_Title, and Org_Group variables for the table. I used a where clause to only include employees whose last names started with W, Y, or Z.

```
proc sql ;
 select Employee_Name, Job_Title, Org_Group
 from stat440.employee_roster
 where Employee_Name like '% W%' or Employee_Name like '% Y%' or
 Employee_Name like '% Z%';
quit;
```

The employees are all in very different groups all around. Many of them have a sales representative position.

Employee Name	Job Title	Group
Tom Zhou	Sales Manager	Sales Management
Fang Wilson	Sales Rep. II	Shoes
Michael Zubak	Sales Rep. III	Outdoors

Employee Name	Job Title	Group
Samantha Waal	Sales Rep. I	Racket Sports
Matsuoka Wills	Sales Rep. III	Team Sports
Kenisha Winge	Concession Manager	Concession Management
Theresa Weisbarth	Logistics Coordinator I	External
Odudu Zisek	Shipping Manager	Stock & Shipping Americas
Suzon Woyach	Warehouse Assistant I	Stock & Shipping Americas
Robert Whitlock	Marketing Assistant II	Organization
Connie Woods	Accountant I	Accounting & Budgeting
Steven Worton	Auditor I	Auditing & Wages
Kimberly Walcott	HR Manager	Staff Administration
Ahmed Zied	IS Administrator III	IS operations
Robert Walker	BI Administrator IV	IS operations
Johannes Wade	Office Assistant IV	Administration
Troyce Van Der Wiele	Warehouse Assistant I	Stock Admin
Michael Westlund	Sales Rep. II	Assorted Sports Articles
Donald Washington	Sales Rep. II	Clothes
Jaime Wetherington	Sales Rep. II	Clothes
Lionel Wende	Sales Rep. I	Racket Sports
Okema Whipkey	Temp. Sales Rep.	Temporary

Problem 1c

I selected the variables Employee_Name, Job_Title, Department, and Employee_Term_Date and used a where clause to make sure that the only observations shown would be those who were terminated after 2006.

```
proc sql;
 select Employee_Name, Job_Title, Department, Employee_Term_Date
 from stat440.employee_roster
 where Employee_Term_Date >= '01JAN2007'd;
quit;
```

Every person in this table is from the sales department. They were all terminated in the first half of the year of 2007. Most of them were Temporary Sales Representatives so it makes sense as to why they were terminated.

Employee Name	Job Title	Department	Employee Termination Date
Lorian Cantatore	Temp. Sales Rep.	Sales	31MAR2007
Ari Moore	Temp. Sales Rep.	Sales	30APR2007
Sharon Bahlman	Temp. Sales Rep.	Sales	31JAN2007
Merryn Quinby	Temp. Sales Rep.	Sales	30JUN2007
Reyne Catenacci	Temp. Sales Rep.	Sales	31MAY2007
Mihailo Lachlan	Temp. Sales Rep.	Sales	31MAR2007
David Anstey	Sales Rep. I	Sales	01FEB2007
Roger Mandzak	Sales Rep. I	Sales	01JAN2007
Karen Grzebien	Sales Rep. I	Sales	01MAR2007
Lawrie Clark	Sales Rep. I	Sales	01MAR2007
Libby Levi	Sales Rep. I	Sales	01MAY2007
Salim Maholo	Temp. Sales Rep.	Sales	30APR2007
Karen Costine	Temp. Sales Rep.	Sales	31JAN2007
Okema Whipkey	Temp. Sales Rep.	Sales	30JUN2007
Aquilla O'Carroll	Temp. Sales Rep.	Sales	30APR2007
Mary Bond-Teague	Temp. Sales Rep.	Sales	31MAR2007
Paul Lawson	Temp. Sales Rep.	Sales	31MAY2007

Employee Name	Job Title	Department	Employee Termination Date
Bruce Armogida	Temp. Sales Rep.	Sales	28FEB2007
Susan Labach	Temp. Sales Rep.	Sales	28FEB2007
Halouise Cassone	Temp. Sales Rep.	Sales	31MAY2007
Stancey Scarbrough	Temp. Sales Rep.	Sales	31MAR2007
Randy Helyar	Temp. Sales Rep.	Sales	31MAR2007

Problem 1d

I used create table so that I could create a temporary data set called lowEarners_hmkang3. I selected Employee_Namr, Job_Title, and Salary. Next to Job_Title and Salary, I labelled Position and Yearly Salary respectively. I used a where clause to indicate that the table should only include those who had a salary that was less than $30,000. I used a select clause to print the data portion and a where clause in order to only get employees whose name starts with ‘a’.

```
proc sql ;
 create table lowEarners_hmkang3 as
 select Employee_Name, Job_Title 'Position', Salary 'Yearly
 Salary'
 from stat440.employee_roster
 where Salary<30000;
 select Employee_Name, Job_Title, Salary
 from lowEarners_hmkang3
 where Employee_Name like 'A%';
quit;
```

Most of the positions for the employees are sales representatives or assistant positions.

Employee Name	Position	Yearly Salary
Austen Ralston	Service Assistant II	$29,250
Atul Leyden	Sales Rep. I	$26,605
Amanda Liebman	Sales Rep. II	$27,465
Alban Kingston	Sales Rep. III	$28,830

Employee Name	Position	Yearly Salary
Alena Moody	Sales Rep. II	$26,205
Andrew Conolly	Sales Rep. I	$25,745
Ari Moore	Temp. Sales Rep.	$25,820
Anthony Nichollas	Trainee	$26,185
Angela Landry	Concession Assistant I	$26,840
Anita Howard	Warehouse Assistant I	$25,130
Angelia Neal	Marketing Assistant II	$28,535
Allan Rudder	Sales Rep. II	$26,165
Asishana Polky	Sales Rep. I	$25,110
Agnieszka Holthouse	Sales Rep. III	$29,385
Azmi Mees	Sales Rep. I	$25,770
Albert Knapp	Temp. Sales Rep.	$26,370
Aquilla O'Carroll	Temp. Sales Rep.	$26,430

Problem 2a

I used separate proc sql steps to create temporary sas data sets discountRet_hmkang3, discountCat_hmkang3, and discountInt_hmkang3. I selected Customer_ID, Customer_Name, and Total_Sales which was a sum of the Total_Retail_Price for each individual. I used a where clause so that each data set would only contain the order types that match with their name. I grouped by customer_ID and customer_name so that there would not be several of the same customer. I made sure that the total_sales was greater than 200 since we only want to reward the customers that spent more than $200.

```
proc sql;
 create table discountRet_hmkang3 as
 select Customer_ID, Customer_Name, sum(Total_Retail_Price) as
 Total_Sales format=dollar9.2
 from stat440.orders2018
```

```
 where Order_Type=1
 group by Customer_ID, Customer_Name
 having sum(Total_Retail_Price)>200;
 describe table discountRet_hmkang3;
quit;
```

The log showed the descriptor portion and described the process and format of each dataset.

```
NOTE: SQL table WORK.DISCOUNTRET_HMKANG3 was created like:

 create table WORK.DISCOUNTRET_HMKANG3( bufsize=131072 )
   (
    Customer_ID num format=12. label='Customer ID',
    Customer_Name char(40) label='Customer Name',
    Total_Sales num format=DOLLAR9.2
   );


proc sql;
 create table discountCat_hmkang3 as
 select Customer_ID, Customer_Name, sum(Total_Retail_Price) as
 Total_Sales format=dollar9.2
 from stat440.orders2018
 where Order_Type=2
 group by Customer_ID, Customer_Name
 having sum(Total_Retail_Price)>200;
 describe table discountCat_hmkang3;
quit;
NOTE: SQL table WORK.DISCOUNTCAT_HMKANG3 was created like:

 create table WORK.DISCOUNTCAT_HMKANG3( bufsize=131072 )
   (
    Customer_ID num format=12. label='Customer ID',
    Customer_Name char(40) label='Customer Name',
    Total_Sales num format=DOLLAR9.2
   );


proc sql;
 create table discountInt_hmkang3 as
 select Customer_ID, Customer_Name, sum(Total_Retail_Price) as
 Total_Sales format=dollar9.2
 from stat440.orders2018
 where Order_Type=3
 group by Customer_ID, Customer_Name
 having sum(Total_Retail_Price)>200;
 describe table discountInt_hmkang3;
quit;
```

```
NOTE: SQL table WORK.DISCOUNTINT_HMKANG3 was created like:

 create table WORK.DISCOUNTINT_HMKANG3( bufsize=131072 )
   (
    Customer_ID num format=12. label='Customer ID',
    Customer_Name char(40) label='Customer Name',
    Total_Sales num format=DOLLAR9.2
   );
```

Problem 2b

I titled the data sets ‘retail sales’, ‘catalog sales’, and ‘internet sales’ since they were appropriate names that described the data sets well.

```
proc sql;
 title 'Retail Sales';
 select *
 from discountRet_hmkang3;
quit;
```

There were several customers in retail sales and internet sales but only one in catalog sales for the customers who paid over $200.

Retail Sales

Customer ID	Customer Name	Total_Sales
5	Sandrina Stephano	$213.10
10	Karen Ballinger	$3,479.09
12	David Black	$253.20
31	Cynthia Martinez	$1,322.30
45	Dianne Patchin	$700.28
89	Wynella Lewis	$602.00
90	Kyndal Hooks	$221.70

```
proc sql;
 title 'Catalog Sales';
 select *
 from discountCat_hmkang3;
quit;
```

Catalog Sales

Customer ID	Customer Name	Total_Sales
90	Kyndal Hooks	$319.00

```
proc sql;
 title 'Internet Sales';
 select *
 from discountInt_hmkang3;
quit;
```

Internet Sales

Customer ID	Customer Name	Total_Sales
5	Sandrina Stephano	$478.00
24	Robyn Klem	$358.80
27	Cynthia Mccluney	$1,093.60
31	Cynthia Martinez	$455.30
34	Alvan Goheen	$642.50

Problem 2c

I used as in the from clause so that I wouldn't have to type out the whole data set name each time I wanted to refer to a variable within that dataset. I used a where clause so that I could find which customers were in 2 or more of the order type data sets.

```
proc sql;
 create table silver_hmkang3 as
 select *
 from discountRet_hmkang3 as dr,discountCat_hmkang3 as
 dc,discountInt_hmkang3 as di
 where dr.customer_ID=dc.customer_ID or
 di.customer_ID=dr.customer_ID or di.customer_ID=dc.customer_ID;
 describe table silver_hmkang3;
quit;
```

I printed the descriptor portion which showed the process and format.

```
NOTE: SQL table WORK.SILVER_HMKANG3 was created like:

 create table WORK.SILVER_HMKANG3( bufsize=131072 )
```

```
   (
    Customer_ID num format=12. label='Customer ID',
    Customer_Name char(40) label='Customer Name',
    Total_Sales num format=DOLLAR9.2
   );

 80         quit;
```

Problem 2d

I printed the data portion with proc sql by selecting variables from the dataset.

```
proc sql;
 select *
 from silver_hmkang3;
quit;
```

There were only 3 people that spend $200 or more for 2 or more order types.

Customer ID	Customer Name	Total_Sales
5	Sandrina Stephano	$213.10
31	Cynthia Martinez	$1,322.30
90	Kyndal Hooks	$221.70
90	Kyndal Hooks	$221.70
90	Kyndal Hooks	$221.70
90	Kyndal Hooks	$221.70
90	Kyndal Hooks	$221.70

Problem 2e

I used a similar method to the silver_hmkang3 data set except I had to adjust it for the fact that the customer needed more than $350 spent on 2 or more order types.

```
proc sql;
 create table gold_hmkang3 as
 select *
 from discountRet_hmkang3 as dr,discountCat_hmkang3 as
 dc,discountInt_hmkang3 as di
 where (dr.Total_Sales>=350 and di.Total_Sales>=350 or
 dc.Total_Sales>=350)
 and (dr.customer_ID=dc.customer_ID or
 di.customer_ID=dr.customer_ID or
```

```
 di.customer_ID=dc.customer_ID);
 describe table gold_hmkang3;
quit;
```

I printed the descriptor portion which showed the process and the formatting done.

```
NOTE: SQL table WORK.GOLD_HMKANG3 was created like:

 create table WORK.GOLD_HMKANG3( bufsize=131072 )
   (
    Customer_ID num format=12. label='Customer ID',
    Customer_Name char(40) label='Customer Name',
    Total_Sales num format=DOLLAR9.2
   );

 81         quit;
```

Problem 2f

I printed the data portion by selecting all of the variables in the proc sql step.

```
proc sql;
 select *
 from gold_hmkang3;
quit;
```

There was only one person that had more than $350 spent in 2 or more order types.

Customer ID	Customer Name	Total_Sales
31	Cynthia Martinez	$1,322.30