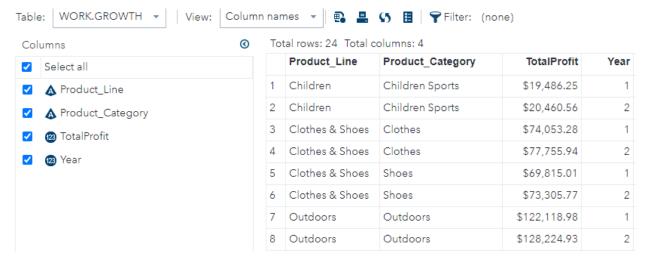
Preparing for the SAS Programming Certification

Week 3: Controlling DATA Step

- 1. Processing Data Review
- 2. Summarizing Data Review
- 3. Manipulating Data with Function Review
- 4. Creating and Using Custom Format Review



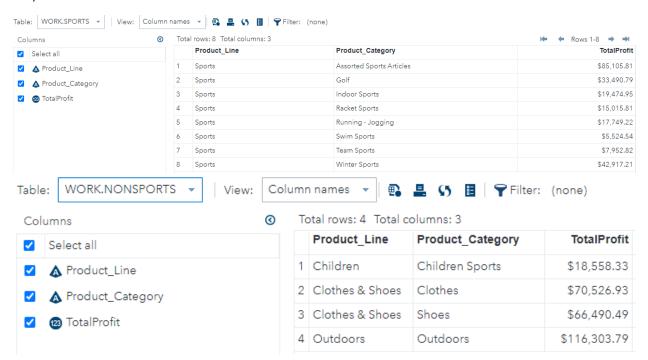
data sports nonsports;

set cr.profit_summary;

if Product Line="Sports" then output sports;

else output nonsports;

run;



/*Programming Question 4.01

If necessary, start SAS Studio and submit libname.sas.

Open p201q1.sas from the programs folder. Fix the program to ensure that the following actions occur:

Values of Country should all be uppercase in the output tables.

Include only rows where TermDate is missing.

All rows where Country is US are written to the emp_US table.

All rows where Country is AU are written to the emp_AU table.

How many rows are in the emp_US table?

How many rows are in the emp_AU table?

*/

data emp_US emp_AU;

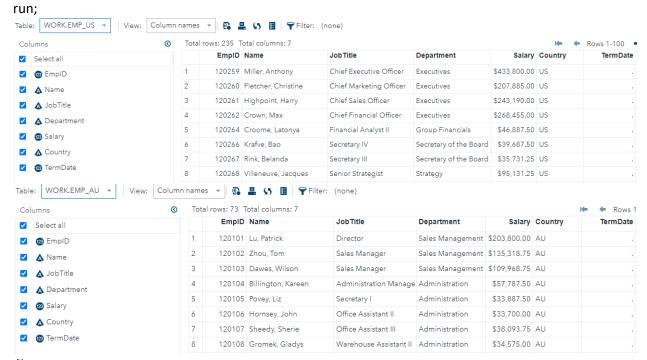
set cr.employee(keep=EmpID Name JobTitle Salary Department Country TermDate);

where TermDate=.;

Country=upcase(country);

if Country="US" then output emp_US;

else output emp_AU;



/*Programming Question 4.02

If necessary, start SAS Studio.

Write a new program that reads the sashelp.heart table and separates rows into two tables, dead and alive, based on the value of Status.

Drop Status from both tables and drop DeathCause and AgeAtDeath from the alive table.

How many rows are in the dead table?

What is the value of Cholesterol for row 100 in the alive table?

How many columns are in the alive table?

*/

```
data dead alive;

set sashelp.heart;

if Status="Dead" then do;

drop Status;

output dead;

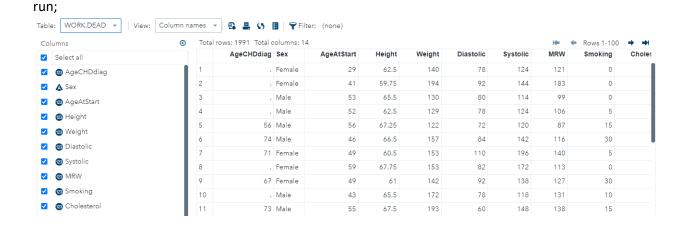
end;

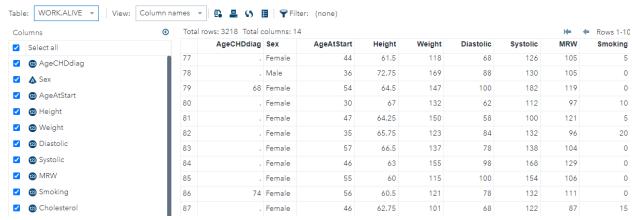
else if Status="Alive" then do;

drop Status DeathCause AgeAtDeath;

output alive;

end;
```





/*Programming Question 4.03

If necessary, start SAS Studio and submit libname.sas. Write a new program to do the following:

Read the cr.employee_current table and create a summary table named salary that calculates TotalSalary (sum of Salary) for each department.

In the output table, include only one row for each department.

Create a new table named salaryforecast that reads the salary summary table.

For each row in the salary table, write three rows to the salaryforecast table.

For the first row written to the salaryforecast table, create a column named Year that is equal to 1.

Increase TotalSalary by 3% (multiply by 1.03).

Write additional rows to the output table with Year equal to 2 and 3 and TotalSalary increased by an additional 3%.

Display TotalSalary as currency values rounded to the nearest whole number.

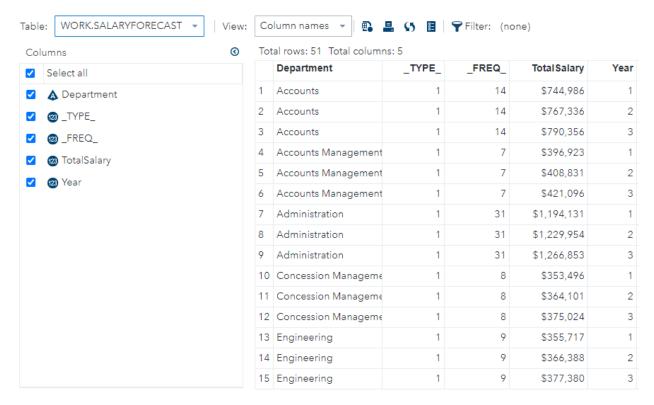
What is the total salary for the Strategy Department before the forecast is performed?

Note: Round your answer to the nearest whole number and enter only numbers.

How many rows are in the salaryforecast table?

```
*/
proc means data=cr.employee_current noprint;
  var Salary;
  class Department;
  output out=salary sum=TotalSalary;
  ways 1;
```

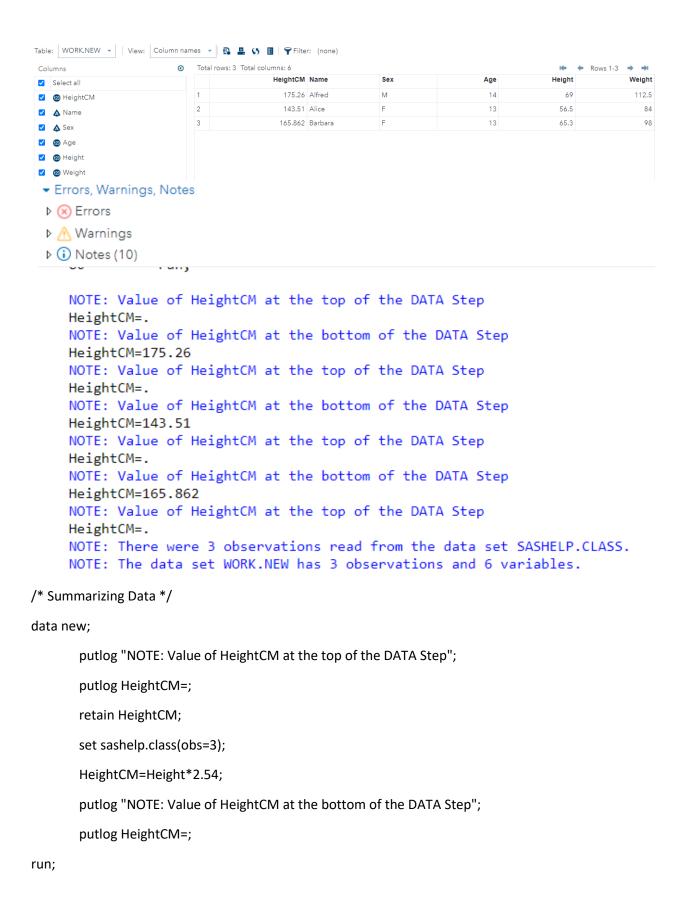
data salaryforecast; set salary; format TotalSalary dollar12.; Year=1; TotalSalary=TotalSalary*1.03; output; Year=2; TotalSalary=TotalSalary*1.03; output; Year=3; TotalSalary=TotalSalary*1.03; output; run; WORK.SALARY Column names ▼ | 🖺 🚨 👣 🖺 | 👕 Filter: (none) Table: View: Total rows: 17 Total columns: 4 **③** Columns Department _TYPE_ FREQ **Total Salary** Select all Accounts \$723,287.50 1 14 **▲** Department Accounts Management 7 \$385,362.50 23 _TYPE_ 3 Administration 31 \$1,159,350.00 23 _FREQ_ \$343,200.00 4 Concession Manageme 1 8 TotalSalary 5 Engineering 9 \$345,356.25 6 Executives 4 \$1,153,330.00 1 Group Financials 1 1 \$46,887.50 Group HR Managemen 15 \$764,131.25 9 20 \$1,265,062,50 10 Logistics Management \$836,025.00 1 12 11 Marketing 19 \$1,038,906.25 1 12 Purchasing \$739,068.75 1 15 122 \$4,246,800.00 13 Sales 1 14 Sales Management 10 \$1,289,425.00 1 15 Secretary of the Board \$75,418.75 1 2 16 Stock & Shipping 18 \$788,537.50 Property Value 17 Strategy \$95,131.25 Label



/* Summarizing Data */

data new;

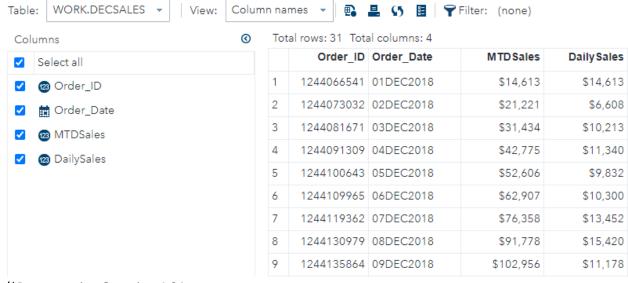
```
putlog "NOTE: Value of HeightCM at the top of the DATA Step";
       putlog HeightCM=;
       set sashelp.class(obs=3);
        HeightCM=Height*2.54;
       putlog "NOTE: Value of HeightCM at the bottom of the DATA Step";
       putlog HeightCM=;
run;
```



putlog "NOTE: Value of HeightCM at the bottom of the DATA Step";

```
putlog HeightCM=;
run;
▼ Errors, Warnings, Notes
 ▶ (x) Errors
 ▶ Marnings
 72
    73
               /* Summarizing Data */
    74
               data new;
               putlog "NOTE: Value of HeightCM at the top of the DATA Step";
    75
    76
               putlog HeightCM=;
               retain HeightCM 0;
    77
               set sashelp.class(obs=3);
    78
    79
               HeightCM=Height*2.54;
               putlog "NOTE: Value of HeightCM at the bottom of the DATA Step";
    80
    81
               putlog HeightCM=;
    82
               run;
    NOTE: Value of HeightCM at the top of the DATA Step
   HeightCM=0
    NOTE: Value of HeightCM at the bottom of the DATA Step
   HeightCM=175.26
    NOTE: Value of HeightCM at the top of the DATA Step
   HeightCM=175.26
    NOTE: Value of HeightCM at the bottom of the DATA Step
   HeightCM=143.51
    NOTE: Value of HeightCM at the top of the DATA Step
   HeightCM=143.51
    NOTE: Value of HeightCM at the bottom of the DATA Step
   HeightCM=165.862
    NOTE: Value of HeightCM at the top of the DATA Step
    HeightCM=165.862
    NOTE: There were 3 observations read from the data set SASHELP.CLASS.
    NOTE: The data set WORK.NEW has 3 observations and 6 variables.
proc sort data=cr.profit out=decDaily;
      where month(Order_Date)=12;
      by Order Date;
run;
data DecSales:
      set decDaily;
      retain MTDSales 0;
```

```
MTDSales=sum(MTDSales,Profit);
       format MTDSales dollar12.;
       keep Order_ID Order_Date Profit MTDSales;
run;
Table:
                             View:
                                    Column names 🔻
                                                      🖺 💄 😘 📱 🕆 Filter: (none)
                                     ③
                                          Total rows: 1200 Total columns: 4
 Columns
                                                                                       MTDSales
                                                  Order_ID Order_Date
                                                                             Profit
 Select all
                                          1
                                                1244072351 01DEC2018
                                                                           $142.55
                                                                                           $143
 ✓ Ø Order_ID
                                          2
                                                                                           $542
                                                1244072328 01DEC2018
                                                                           $399.00
     $593
                                          3
                                                1244072170 01DEC2018
                                                                            $51.80
     Profit
                                          4
                                                1244072131 01DEC2018
                                                                           $139.05
                                                                                           $732
 MTDSales
                                          5
                                                1244071500 01DEC2018
                                                                           $130.40
                                                                                           $863
                                                1244071345 01DEC2018
                                                                            $14.75
                                                                                           $878
                                          6
                                          7
                                                1244071081 01DEC2018
                                                                            $67.04
                                                                                           $945
                                          8
                                                1244070912 01DEC2018
                                                                           $366.80
                                                                                          $1,311
proc sort data=cr.profit out=decDaily;
       where month(Order_Date)=12;
       by Order_Date;
run;
data DecSales;
       set decDaily;
       by Order_Date;
       MTDSales+Profit;
       if first.Order_Date=1 then DailySales=0;
       DailySales+Profit;
       if last.Order_Date=1;
       format MTDSales DailySales dollar12.;
       keep Order_ID Order_Date MTDSales DailySales;
run;
```



/*Programming Question 4.04

If necessary, start SAS Studio. Write a new program to do the following:

Read the sashelp.stocks table and create a new table named stocks_total.

Include only rows where the Date is in 2005.

Create an accumulating column for Volume named YTDVolume that is a running total of Volume for each stock.

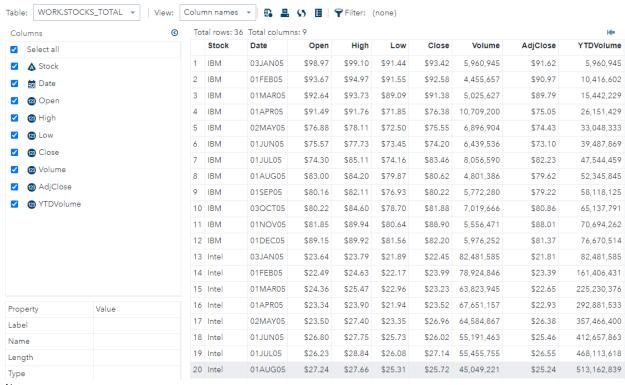
How many rows are in the stocks_total table?

What is the value for YTDVolume for Intel on 01AUG2005?

```
*/
```

format YTDVolume comma16.;

run;



/*Programming Question 4.05

If necessary, start SAS Studio. Write a new program to do the following:

Read the sashelp.shoes table and create a new table named highlow.

The highlow table should include one row for the lowest value of Sales for each Product value,

and another row for the highest value of Sales for each Product value.

Create a column named HighLow that includes the value High or Low depending on the value of Sales.

How many rows are in the highlow table?

Which subsidiary has the highest Sales value for Sandals?

Which region has the lowest Sales value for Boot?

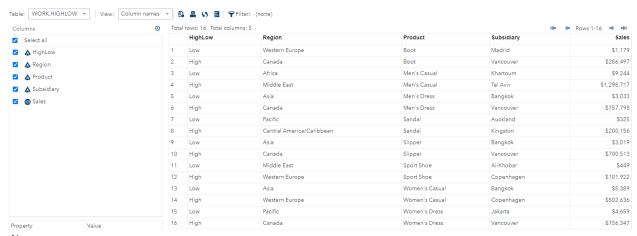
*/

proc sort data=sashelp.shoes out=highlow;

by Product Sales;

```
data highlow;
length HighLow $ 4;
set highlow;
by Product;
if first.product then do;
    HighLow="Low";
    output;
end;
if last.product then do;
    HighLow="High";
    output;
end;
keep Region Product HighLow Sales Subsidiary;
```

run;



/*Programming Question 4.06

If necessary, start SAS Studio and submit libname.sas. Open p202q3.sas from the programs folder. Fix the program to ensure that the following actions occur:

There should be one row for each department in the dept salary table.

The TotalDeptSalary column should be the total of Salary within each department.

The LowSalaryJob column should be the job title for the minimum salary paid within each department.

The HighSalaryJob column should be the job title for the maximum salary paid within each department.

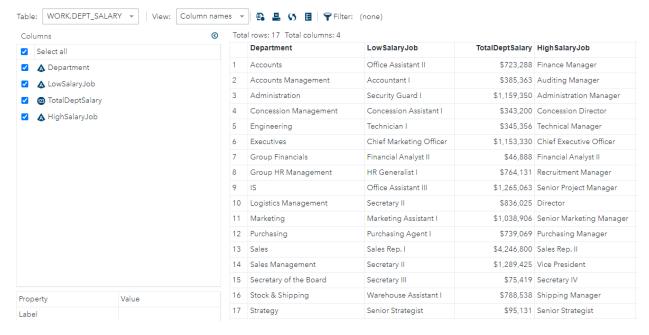
What is the value of TotalDeptSalary for the Accounts Department? Note: enter the numeric value with no formatting.

What is the value of LowSalaryJob for the Engineering Department?

What is the value of HighSalaryJob for the Logistics Management Department?

*/

proc sort data=cr.employee_current out=emp_sort; by Department Salary; run; data dept_salary; set emp_sort; retain LowSalaryJob; by Department; if first.Department then do; TotalDeptSalary=0; LowSalaryJob=JobTitle; end; TotalDeptSalary+Salary; if last.department then do; HighSalaryJob=JobTitle; output; end; keep Department TotalDeptSalary HighSalaryJob LowSalaryJob; format TotalDeptSalary dollar12.;



/* Manipulating Data with Function */

data qtr_detail;

set cr.qtr_sales;

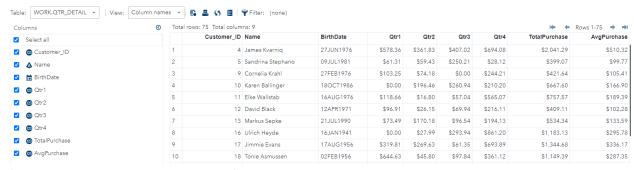
TotalPurchase=sum(of qtr:);

*TotalPurchase=sum(qtr1, qtr2, qtr3, qtr4);

AvgPurchase=round(mean(of qtr:), .01);

format TotalPurchase AvgPurchase dollar12.2;

run;



/* Manipulating Data with Function */

data qtr_detail;

set cr.qtr_sales;

TotalPurchase=sum(of qtr:);

*TotalPurchase=sum(qtr1, qtr2, qtr3, qtr4);

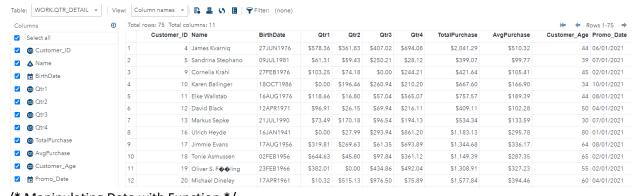
AvgPurchase=round(mean(of qtr:), .01);

Customer_Age=int(yrdif(BirthDate, today(), "age"));

Promo_Date=mdy(month(BirthDate), 1, year(today()));

format TotalPurchase AvgPurchase dollar12.2 Promo_Date mmddyy10.;

run;



/* Manipulating Data with Function */

```
data qtr_detail;
```

set cr.qtr_sales;

TotalPurchase=sum(of qtr:);

*TotalPurchase=sum(qtr1, qtr2, qtr3, qtr4);

AvgPurchase=round(mean(of qtr:), .01);

Customer_Age=int(yrdif(BirthDate, today(), "age"));

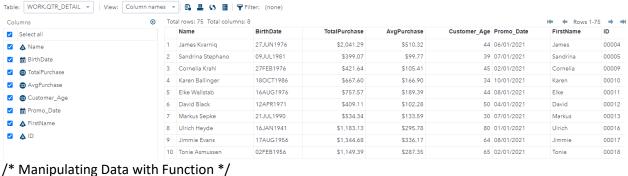
Promo_Date=mdy(month(BirthDate), 1, year(today()));

FirstName=scan(Name, 1, "");

ID=put(Customer ID, z5.);

format TotalPurchase AvgPurchase dollar12.2 Promo Date mmddyy10.;

drop qtr: Customer_ID;

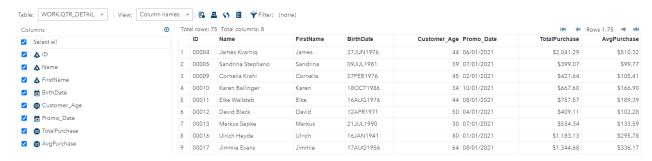


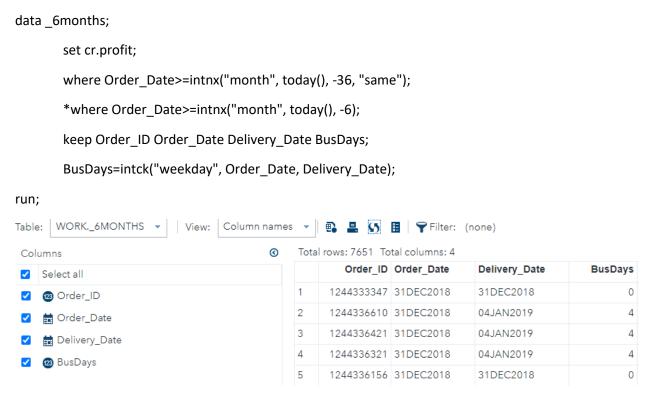
```
data qtr detail;
       set cr.qtr sales;
       TotalPurchase=sum(of qtr:);
       *TotalPurchase=sum(qtr1, qtr2, qtr3, qtr4);
       AvgPurchase=round(mean(of qtr:), .01);
       Customer Age=int(yrdif(BirthDate, today(), "age"));
       Promo_Date=mdy(month(BirthDate), 1, year(today()));
       FirstName=scan(Name, 1, "");
       ID=put(Customer_ID, z5.);
       format TotalPurchase AvgPurchase dollar12.2 Promo_Date mmddyy10.;
       drop qtr: Customer_ID;
```

data qtr_detail;

retain ID Name FirstName BirthDate Customer_Age Promo_Date TotalPurchase AvgPurchase; set qtr_detail;

run;





/* Programming Question 4.07

If necessary, start SAS Studio. Write a new program to do the following:

Read the sashelp.fish table and create a new temporary table named fish.

Compute a new column named Length that is the mean of Length1 through Length3. Round the stored values to the nearest hundredth.

Create a summary report that calculates the average of Length for each value of Species. Round the average to two decimal places.

What is the value of Length for the second row in the fish table?

How many Perch fish were measured?

What is the average length for all Pike fish?

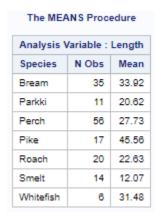
```
*/
data fish;
set sashelp.fish;
Length=round(mean(of Length:), 0.01);
run;
```

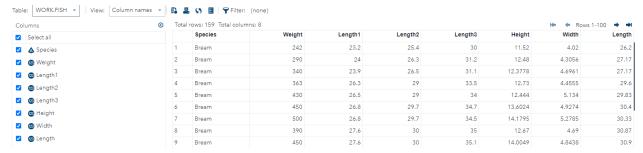
proc means data=fish mean maxdec=2;

var Length;

class Species;

run;





/* Programming Question 4.08

If necessary, start SAS Studio. Write a new program to do the following:

Read the sashelp.baseball table and create a new table named outfield.

Filter rows to include only players where the second letter of Position is F.

Create a new column named Player that rearranges the value in the Name column so that first name is first and then last name separated by a space.

Compute a new column named BatAvg as nHits divided by nAtBat. Round the calculated value to three decimal places.

Sort the outfield table by descending BatAvg.

How many rows are in the outfield table?

What is the value of Player in the first row?

What is the value of BatAvg for Kirby Puckett?

*/

data outfield;

```
set sashelp.baseball;
       where substr(Position, 2, 1)="F";
       Player=catx(" ", scan(Name, 2, ","), scan(Name, 1, ","));
       BatAvg=round(nHits/nAtBat, .001);
run;
proc sort data=outfield out=outfield sort;
       by descending BatAvg;
run;
*Solution;
data outfield;
    set sashelp.baseball;
    where substr(Position, 2, 1)="F";
    Player=catx(" ", scan(Name, 2), scan(Name, 1));
    BatAvg=round(nHits/nAtBat, .001);
    keep Player BatAvg Position;
run;
proc sort data=outfield;
    by descending BatAvg;
run;
Table: WORK.OUTFIELD_SORT ▼ | View: Column names ▼ | 🖺 😃 👣 🗏 | 🍞 Filter: (none)
                      Total rows: 107 Total columns: 26
                                                                                 (← ← Rows 1-100 → →
                          ✓ Select all
                                                                                          BatAvg
Name
                                                                                           0.329
Team
                                                                                         0.324
                                                                                           0.32
nHome
```

/* Programming Question 4.09

If necessary, start SAS Studio and submit libname.sas. Open p203q3.sas from the programs folder.

Fix the program to ensure that the following actions occur:

HireDate is converted to a numeric SAS date value.

Salary is the numeric equivalent of the character column AnnualSalary.

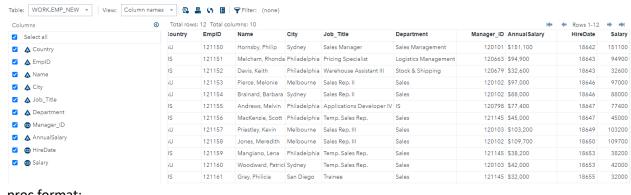
What is the unformatted SAS date value of HireDate for the first row of the emp_new table?

Which informat can be used in the INPUT function to create the Salary column? Note: Do not type the informat width or period.

```
*/
data emp_new;
    set cr.employee_new;
    EmpID=substr(EmpID, 4);
    HireDate2=input(HireDate, anydtdte10.);
    Salary=input(AnnualSalary, dollar12.);
run;

*Solution;
data emp_new;
    set cr.employee_new(rename=(HireDate=HireDateC));
    EmpID=substr(EmpID,4);
    HireDate=input(HireDateC, anydtdte10.);
    Salary=input(AnnualSalary, dollar10.);
    drop HireDateC;
```

run;



proc format;

value shiprange 0="Same day"

1-3="1-3 days"

4-7="4-7 days"

8-high="8+ days"

.="Unknown";

run;

data profit2;

set cr.profit;

format ShipDays shiprange.;

run;



data profit2;

set cr.profit;

ShipRange=put(ShipDays, shiprange.);

run;

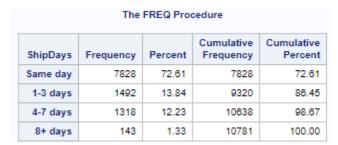


proc freq data=cr.profit;

table ShipDays;

format ShipDays shiprange.;

run;



data country;

```
set cr.country_clean;
          FmtName="$ctryfmt";
          Start=Country_Key;
          Label=Country_Name;
          keep FmtName Start Label;
run;
proc format cntlin=country;
run;
data profit2;
          set cr.profit;
          ShipRange=put(ShipDays, shiprange.);
          format Customer_Country ctryfmt.;
run;
Table: WORK.PROFIT2 ▼ | View: Column names ▼ | 🛍 🚨 👣 🗏 | 😭 Filter: (none)
           ▼ Total rows: 10781 Total columns: 14
Columns
                                                                                                                     r ← Rows 1-100 →
                                 Product_Category Quantity Customer_Country Customer_Continent Order_Source

Assorted Sports Articles 2 Belgium Europe Retail
Outdoors 3 Russia Aria
                                                                                                    Profit ShipDays Age_Range
                                                                                                                             ShipRange
☑ Ø Order_ID
                                                                                                            0 46-60
                                                                                                     $414.60
                                  Asia Internet
Europe Retail
Europe Phone
Europe Retail
Furope Retail
                                                                                                     $946.80
                                                                                                                 4 31-45
                                                                                                                              4-7 days
✓ 🛗 Order_Date
                                                                                                     $322.80
                                                                                                                 4 15-30
                                                                                                                              4-7 days
                                                                                                                  4 15-30
                                                                                                                              4-7 days
0 61-75
                                                                                                                              Same day
✓ ♠ Product_Line
                                                                                                                  0 31-45
                                                                                                                              Same day
```

/* Programming Question 4.10

✓ A Product Category

30 or more

If necessary, start SAS Studio. Write a new program to do the following:

Create a numeric format named BMIRANGE that assigns the following ranges and labels.

Range Label less than 18.5 Underweight 18.5 to 24.9 Normal 25 to 29.9 Overweight

Obese

Create a frequency report based on the BMI column in the sashelp.bmimen table.

Include rows where Age is greater than or equal to 21. Apply the BMIRANGE format to the BMI column.

```
What is the frequency count of men in the Overweight BMI range?

What is the frequency count of men in the Underweight BMI range?

*/

proc format;

value bmirange low-<18.5="Underweight"

18.5-24.9="Normal"

25-29.9="Overweight"

30-high="Obese"

.="Unknown";

run;

proc freq data=sashelp.bmimen;

where Age>=21;

table bmi;

format bmi bmirange.;
```

| | The FREQ Procedure | | | | | | | | | |
|--|--------------------|-----------|---------|-------------------------|-----------------------|--|--|--|--|--|
| | ВМІ | Frequency | Percent | Cumulative Frequency | Cumulative Percent | | | | | |
| | Underweight | 14 | 1.20 | 14 | 1.20 | | | | | |
| | Normal | 318 | 27.25 | 332 | 28.45 | | | | | |
| | Overweight | 521 | 44.64 | 853 | 73.09 | | | | | |
| | Obese | 314 | 26.91 | 1167 | 100.00 | | | | | |

/*Programming Question 4.11

run;

If necessary, start SAS Studio and submit libname.sas. Write a new program to do the following:

Use the cr.continent_codes table to create a numeric format named CONTFMT.

The Code column contains data values, and the Continent column contains labels.

Read the cr.demographics table and compute the sum of the Pop column for each value of Cont.

Apply the CONTFMT format to the Cont column. Note: If you don't have the cr.demograhics table, copy and paste this code into a new program window and run it to create the table.

How many countries are in the European continent?

```
What is the total population in Africa? Note: Type only numbers in your answer.
*/
data continent;
       set cr.continent_codes;
       FmtName="contfmt";
       Start=Code;
       Label=Continent;
       keep FmtName Start Label;
run;
proc format cntlin=continent;
run;
proc means data=cr.demographics sum;
       var Pop;
       class Cont;
       format Pop comma15. Cont contfmt.;
run;
*Solution;
data continentfmt;
  set cr.continent_codes;
  retain fmtname "contfmt";
  Start=Code;
  Label=Continent;
run;
proc format cntlin=continentfmt;
```

run;

proc means data=cr.demographics sum maxdec=0;

var pop;

class cont;

format cont contfmt.;

run;

The MEANS Procedure Analysis Variable: Pop Population (2005) N Obs Numeric Rep. for Continent 509218938 North America 16 377115072 South America 19 Europe 46 779756261 Africa 53 904804386 Asia 4029012970 Oceania 18 72540243.00

/* Programming Question 4.12

If necessary, start SAS Studio and submit libname.sas. Open p204q3.sas from the programs folder.

Fix the program to ensure that the following actions occur:

Create custom formats for the values of Status (M, S, and O) and for ranges of Salary.

Apply the custom formats in the PROC FREQ step.

How many employees are in the Other category?

How many employees in San Diego make between \$50K and \$100K?

*/

proc format;

```
value $statfmt S="Single"

M="Married"

O="Other";

value salrange low-<50000="Under $50K"

50000-100000="50K-100K"

100000<-high="Over 100K";
```

run;

proc freq data=cr.employee;

tables Status;

tables City*Salary / nopercent nocol;

format Status \$statfmt. Salary salrange.;

run;

The FREQ Procedure

| Status | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---------|-----------|---------|-------------------------|-----------------------|
| Married | 215 | 50.71 | 215 | 50.71 |
| Other | 40 | 9.43 | 255 | 60.14 |
| Single | 169 | 39.86 | 424 | 100.00 |

Frequency Row Pct

| Table of City by Salary | | | | | | | | | |
|-------------------------|-------------|-------------|-----------|-------|--|--|--|--|--|
| | Salary | | | | | | | | |
| City(City) | Under \$50K | 50K-100K | Over 100K | Total | | | | | |
| Melbourne | 40 97.56 | 0 0.00 | 1 2.44 | 41 | | | | | |
| Miami-Dade | 70 64.22 | 33 30.28 | 6 5.50 | 109 | | | | | |
| Philadelphia | 67 70.53 | 24 25.26 | 4 4.21 | 95 | | | | | |
| San Diego | 87 77.68 | 20 17.86 | 5 4.46 | 112 | | | | | |
| Sydney | 62 92.54 | 3 4.48 | 2 2.99 | 67 | | | | | |
| Total | 326 | 80 | 18 | 424 | | | | | |