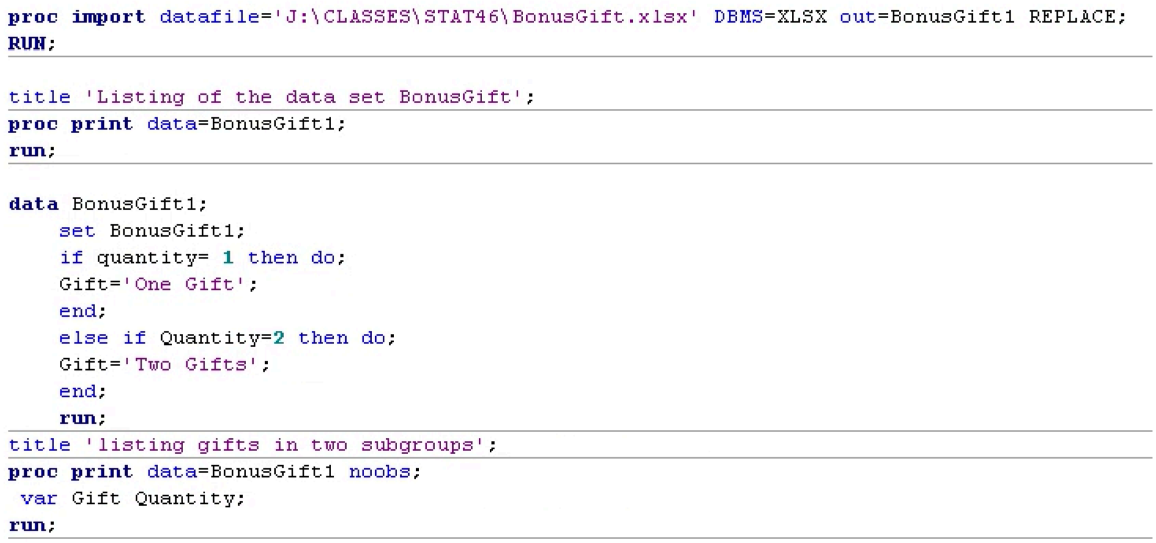
Homework#6 – Solution

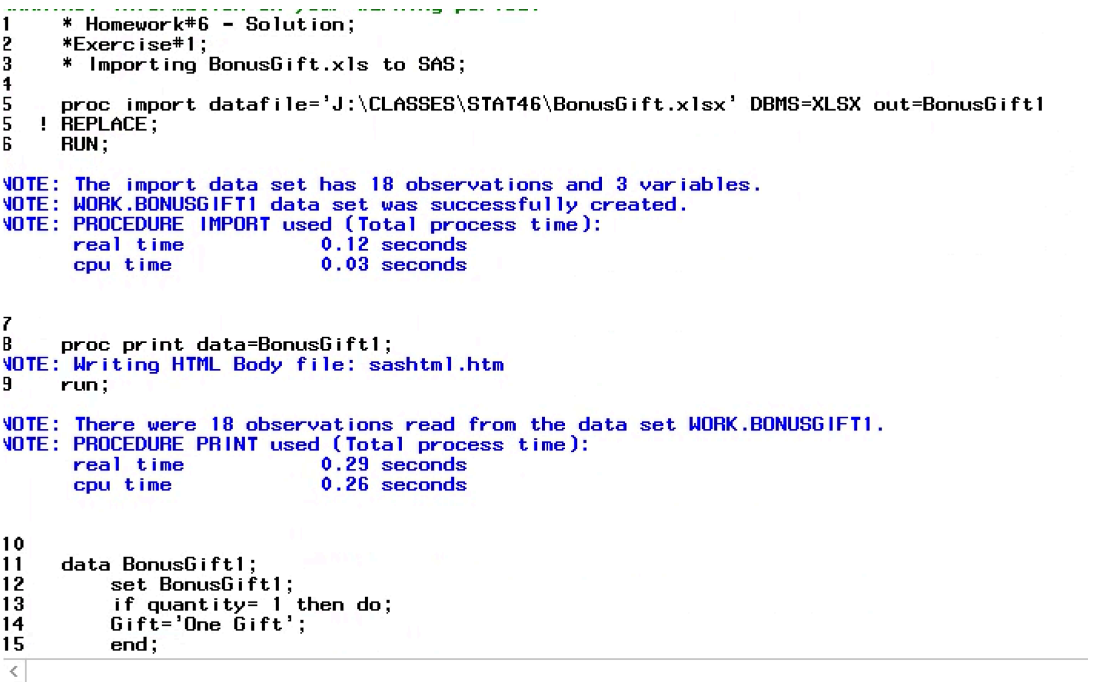
Exercise#1:

Program:

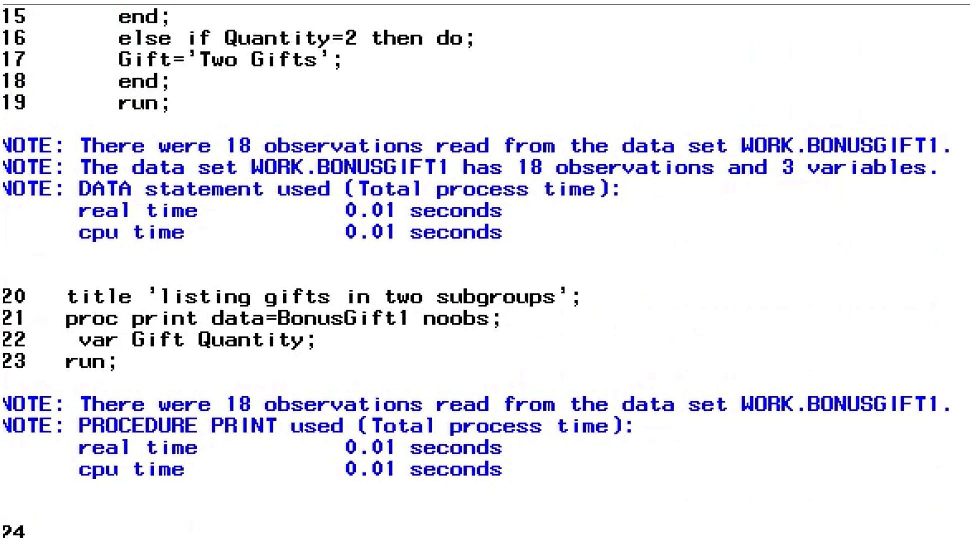


```
proc import datafile='J:\CLASSES\STAT46\BonusGift.xlsx' DBMS=XLSX out=BonusGift1 REPLACE;
RUN;

title 'Listing of the data set BonusGift';
proc print data=BonusGift1;
run;

data BonusGift1;
    set BonusGift1;
    if quantity= 1 then do;
    Gift='One Gift';
    end;
    else if Quantity=2 then do;
    Gift='Two Gifts';
    end;
    run;
title 'listing gifts in two subgroups';
proc print data=BonusGift1 noobs;
 var Gift Quantity;
run;
```

Log:



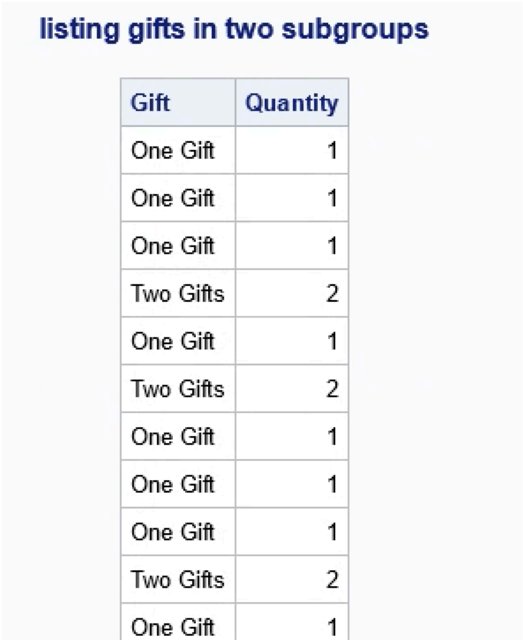
```
1     * Homework#6 - Solution;
2     *Exercise#1;
3     * Importing BonusGift.xls to SAS;
4
5     proc import datafile='J:\CLASSES\STAT46\BonusGift.xlsx' DBMS=XLSX out=BonusGift1
5   ! REPLACE;
6     RUN;

NOTE: The import data set has 18 observations and 3 variables.
NOTE: WORK.BONUSGIFT1 data set was successfully created.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time           0.12 seconds
      cpu time            0.03 seconds


7
8     proc print data=BonusGift1;
NOTE: Writing HTML Body file: sashtml.htm
9     run;

NOTE: There were 18 observations read from the data set WORK.BONUSGIFT1.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.29 seconds
      cpu time            0.26 seconds


10
11    data BonusGift1;
12        set BonusGift1;
13        if quantity= 1 then do;
14        Gift='One Gift';
15        end;
```



```
15       end;
16       else if Quantity=2 then do;
17       Gift='Two Gifts';
18       end;
19       run;

NOTE: There were 18 observations read from the data set WORK.BONUSGIFT1.
NOTE: The data set WORK.BONUSGIFT1 has 18 observations and 3 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


20   title 'listing gifts in two subgroups';
21   proc print data=BonusGift1 noobs;
22    var Gift Quantity;
23   run;

NOTE: There were 18 observations read from the data set WORK.BONUSGIFT1.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


24
```

Output:

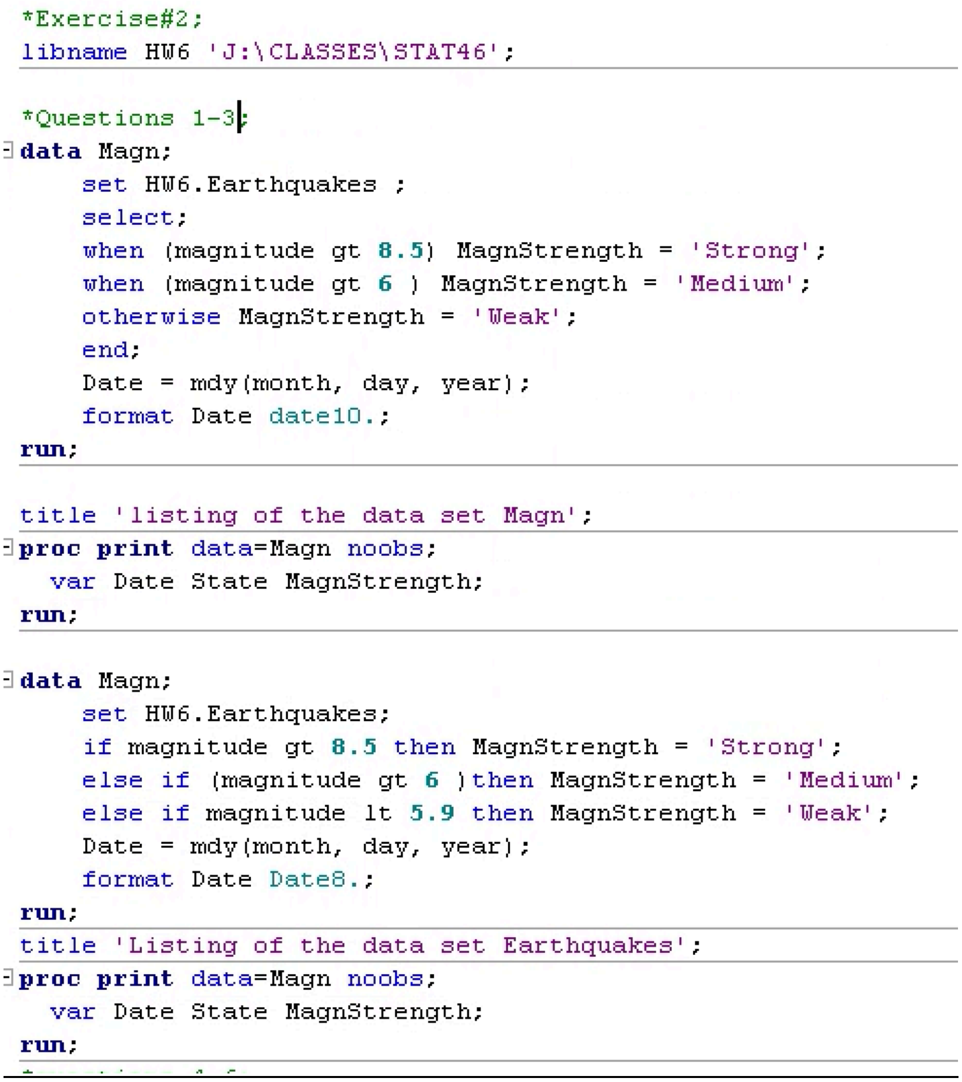
Part of the output table with the 2 subgroups.



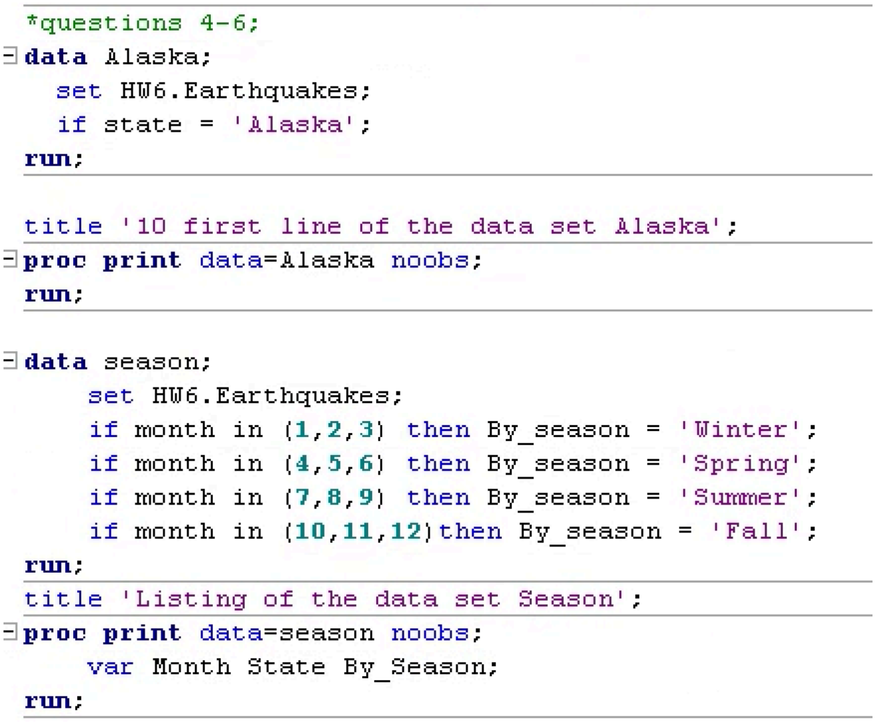
listing gifts in two subgroups

Gift	Quantity
One Gift	1
One Gift	1
One Gift	1
Two Gifts	2
One Gift	1
Two Gifts	2
One Gift	1
One Gift	1
One Gift	1
Two Gifts	2
One Gift	1

Exercise 2:

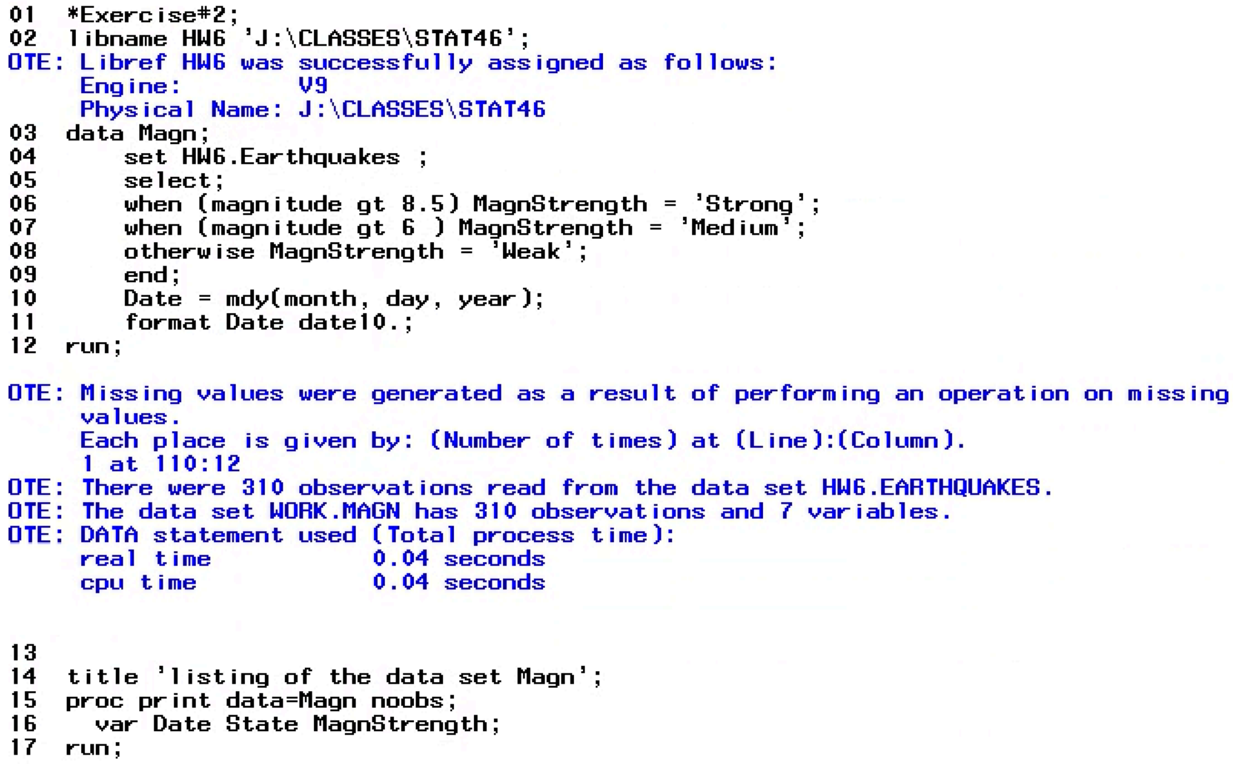


```
*Exercise#2;
libname HW6 'J:\CLASSES\STAT46';

*Questions 1-3;
data Magn;
    set HW6.Earthquakes ;
    select;
    when (magnitude gt 8.5) MagnStrength = 'Strong';
    when (magnitude gt 6 ) MagnStrength = 'Medium';
    otherwise MagnStrength = 'Weak';
    end;
    Date = mdy(month, day, year);
    format Date date10.;
run;

title 'listing of the data set Magn';
proc print data=Magn noobs;
  var Date State MagnStrength;
run;

data Magn;
    set HW6.Earthquakes;
    if magnitude gt 8.5 then MagnStrength = 'Strong';
    else if (magnitude gt 6 )then MagnStrength = 'Medium';
    else if magnitude lt 5.9 then MagnStrength = 'Weak';
    Date = mdy(month, day, year);
    format Date Date8.;
run;
title 'Listing of the data set Earthquakes';
proc print data=Magn noobs;
  var Date State MagnStrength;
run;
```



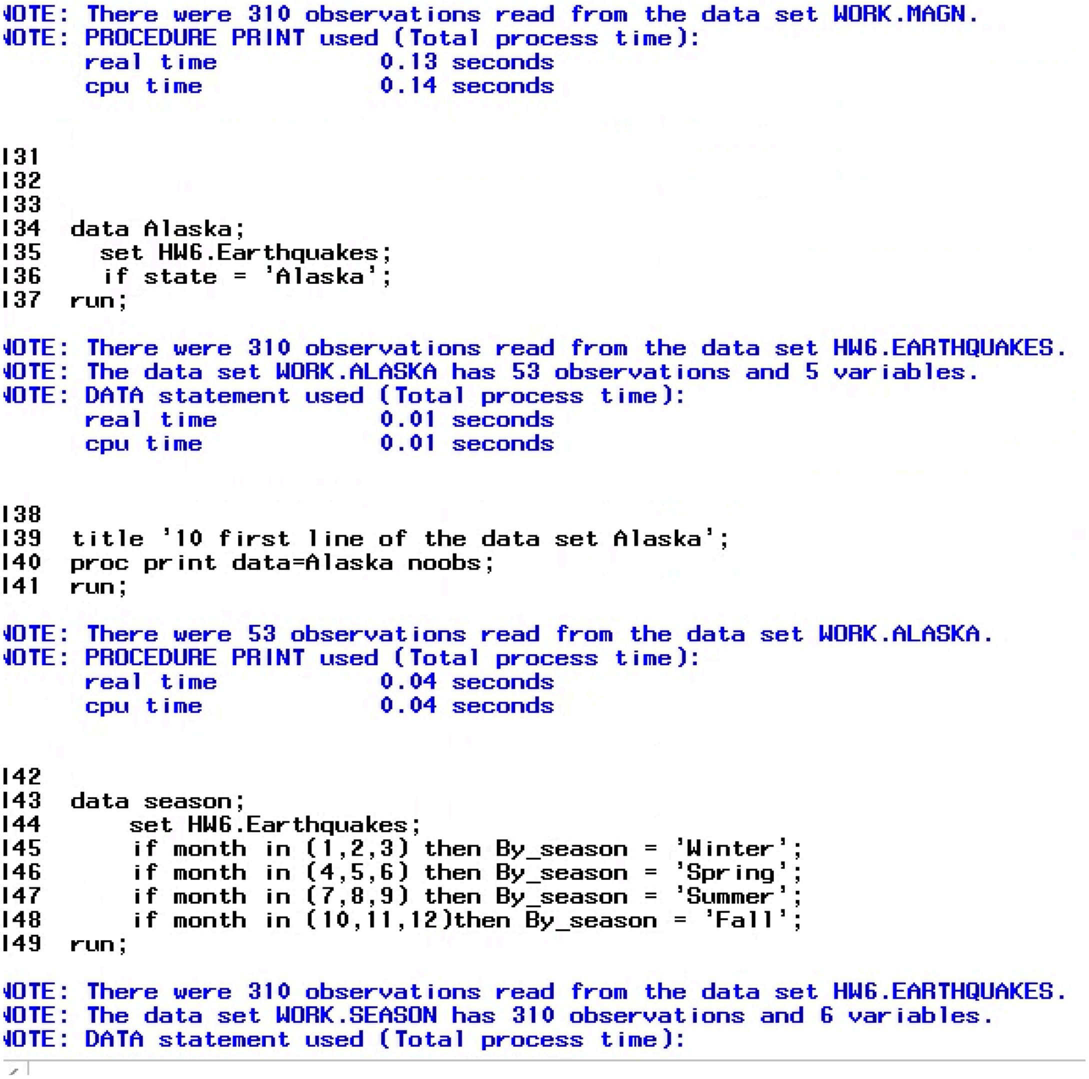
```
*questions 4-6;
data Alaska;
  set HW6.Earthquakes;
  if state = 'Alaska';
run;

title '10 first line of the data set Alaska';
proc print data=Alaska noobs;
run;

data season;
    set HW6.Earthquakes;
    if month in (1,2,3) then By_season = 'Winter';
    if month in (4,5,6) then By_season = 'Spring';
    if month in (7,8,9) then By_season = 'Summer';
    if month in (10,11,12)then By_season = 'Fall';
run;
title 'Listing of the data set Season';
proc print data=season noobs;
    var Month State By_Season;
run;
```

Log:

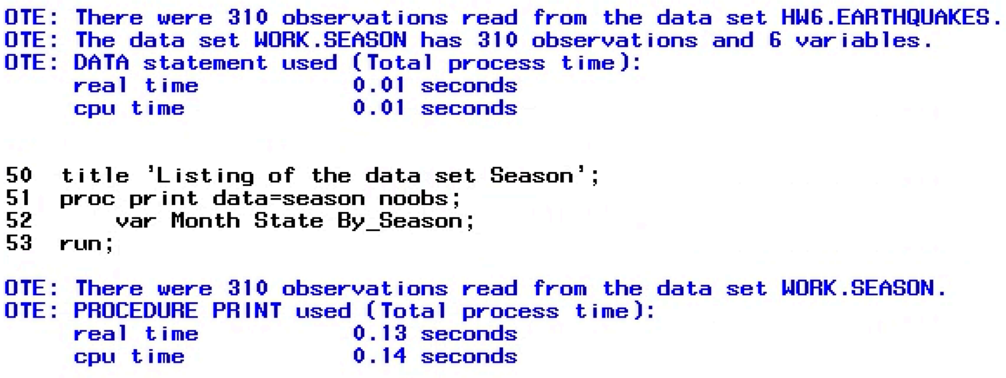


```
01   *Exercise#2;
02   libname HW6 'J:\CLASSES\STAT46';
OTE: Libref HW6 was successfully assigned as follows:
     Engine:        V9
     Physical Name: J:\CLASSES\STAT46
03   data Magn;
04       set HW6.Earthquakes ;
05       select;
06       when (magnitude gt 8.5) MagnStrength = 'Strong';
07       when (magnitude gt 6 ) MagnStrength = 'Medium';
08       otherwise MagnStrength = 'Weak';
09       end;
10       Date = mdy(month, day, year);
11       format Date date10.;
12   run;

OTE: Missing values were generated as a result of performing an operation on missing
     values.
     Each place is given by: (Number of times) at (Line):(Column).
     1 at 110:12
OTE: There were 310 observations read from the data set HW6.EARTHQUAKES.
OTE: The data set WORK.MAGN has 310 observations and 7 variables.
OTE: DATA statement used (Total process time):
     real time           0.04 seconds
     cpu time            0.04 seconds


13
14   title 'listing of the data set Magn';
15   proc print data=Magn noobs;
16     var Date State MagnStrength;
17   run;
```

Log :

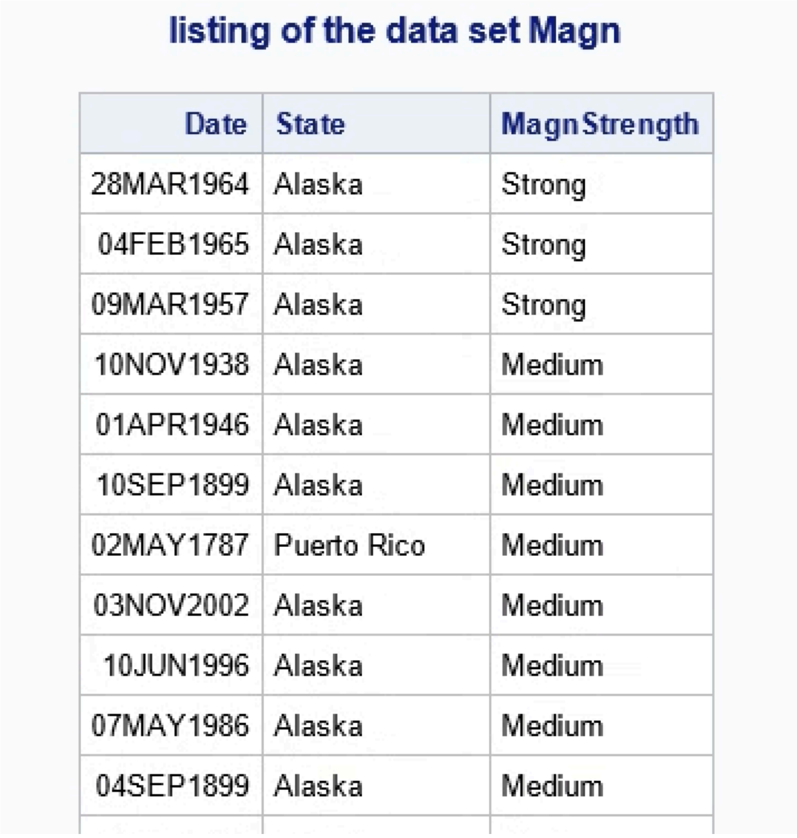


```
NOTE: There were 310 observations read from the data set WORK.MAGN.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.13 seconds
      cpu time            0.14 seconds




     data Alaska;
       set HW6.Earthquakes;
        if state = 'Alaska';
     run;

NOTE: There were 310 observations read from the data set HW6.EARTHQUAKES.
NOTE: The data set WORK.ALASKA has 53 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds



     title '10 first line of the data set Alaska';
     proc print data=Alaska noobs;
     run;

NOTE: There were 53 observations read from the data set WORK.ALASKA.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.04 seconds
      cpu time            0.04 seconds



     data season;
         set HW6.Earthquakes;
         if month in (1,2,3) then By_season = 'Winter';
         if month in (4,5,6) then By_season = 'Spring';
         if month in (7,8,9) then By_season = 'Summer';
         if month in (10,11,12)then By_season = 'Fall';
     run;

NOTE: There were 310 observations read from the data set HW6.EARTHQUAKES.
NOTE: The data set WORK.SEASON has 310 observations and 6 variables.
NOTE: DATA statement used (Total process time):
```



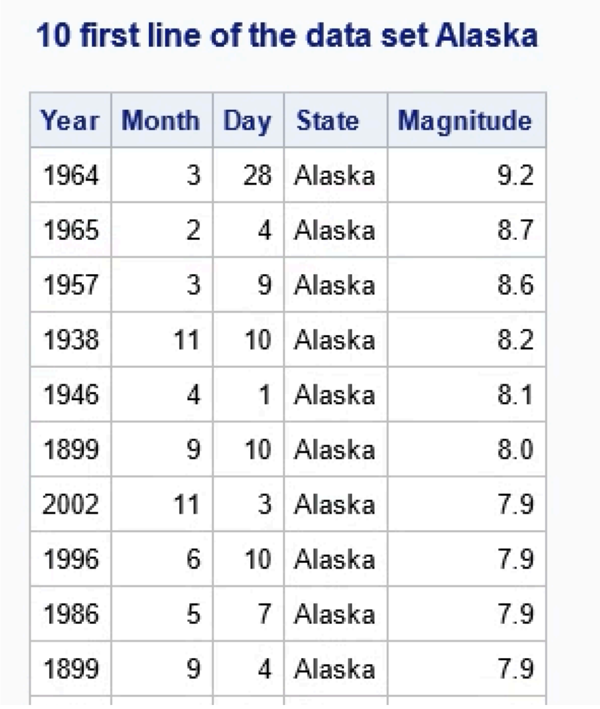
```
OTE: There were 310 observations read from the data set HW6.EARTHQUAKES.
OTE: The data set WORK.SEASON has 310 observations and 6 variables.
OTE: DATA statement used (Total process time):
     real time           0.01 seconds
     cpu time            0.01 seconds


    title 'Listing of the data set Season';
    proc print data=season noobs;
        var Month State By_Season;
    run;

OTE: There were 310 observations read from the data set WORK.SEASON.
OTE: PROCEDURE PRINT used (Total process time):
     real time           0.13 seconds
     cpu time            0.14 seconds
```

Output:



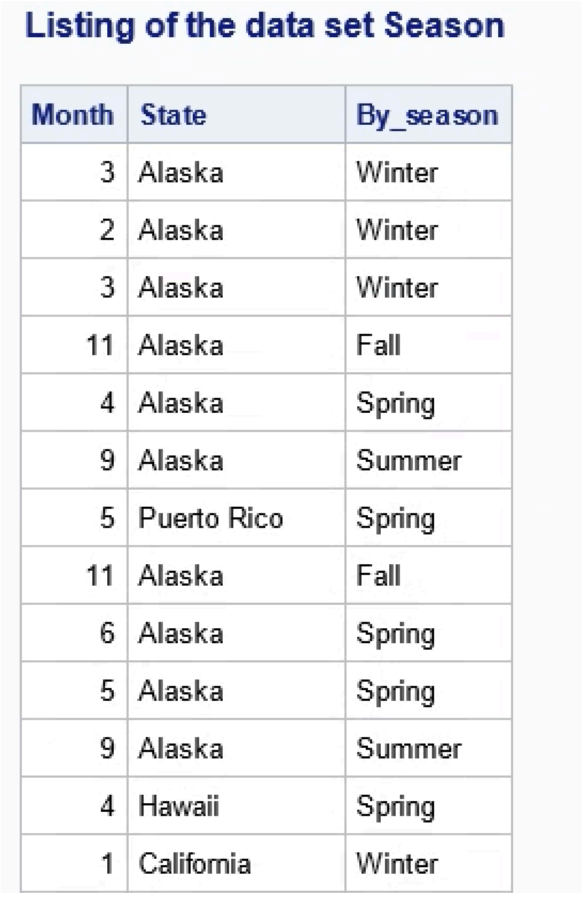
listing of the data set Magn

Date	State	MagnStrength
28MAR1964	Alaska	Strong
04FEB1965	Alaska	Strong
09MAR1957	Alaska	Strong
10NOV1938	Alaska	Medium
01APR1946	Alaska	Medium
10SEP1899	Alaska	Medium
02MAY1787	Puerto Rico	Medium
03NOV2002	Alaska	Medium
10JUN1996	Alaska	Medium
07MAY1986	Alaska	Medium
04SEP1899	Alaska	Medium



10 first line of the data set Alaska

Year	Month	Day	State	Magnitude
1964	3	28	Alaska	9.2
1965	2	4	Alaska	8.7
1957	3	9	Alaska	8.6
1938	11	10	Alaska	8.2
1946	4	1	Alaska	8.1
1899	9	10	Alaska	8.0
2002	11	3	Alaska	7.9
1996	6	10	Alaska	7.9
1986	5	7	Alaska	7.9
1899	9	4	Alaska	7.9



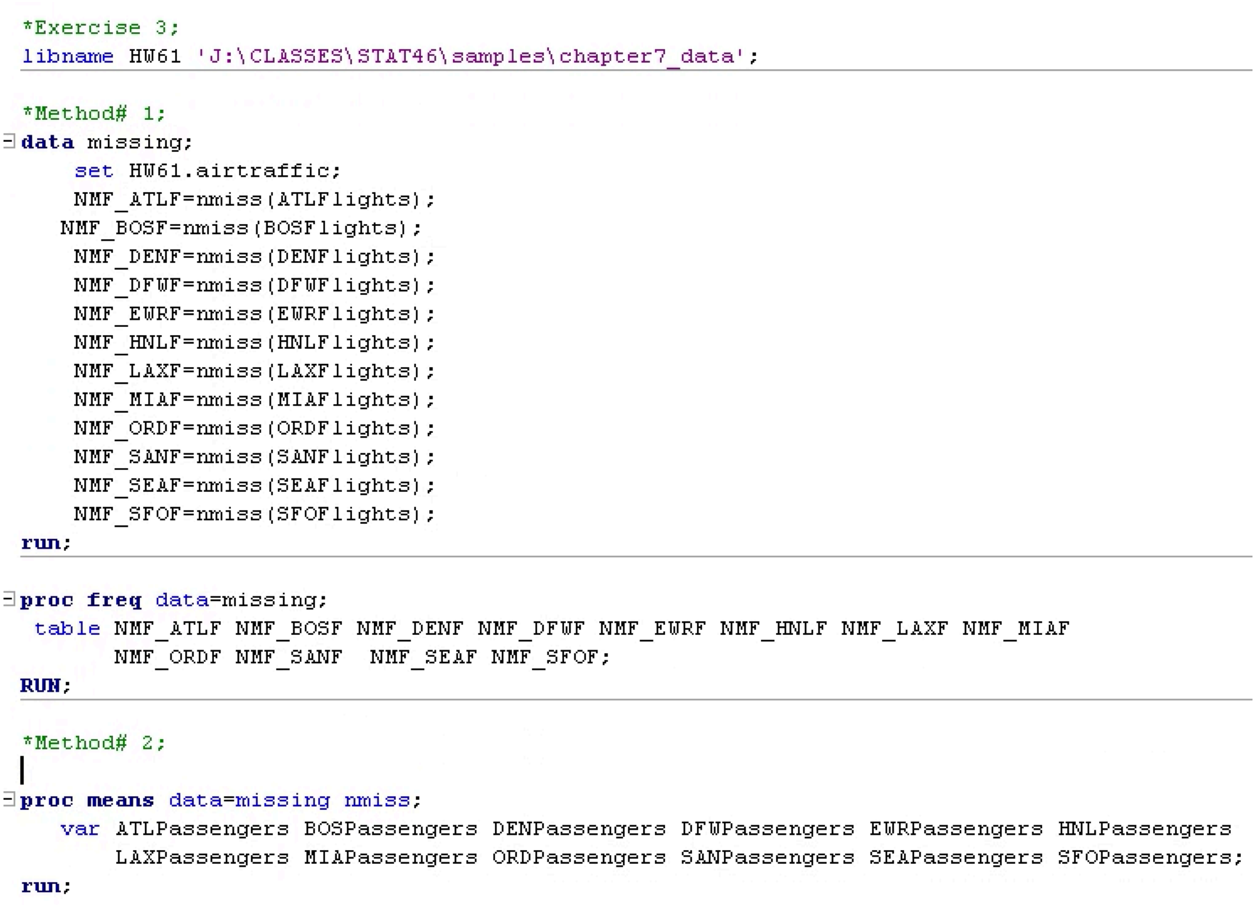
Listing of the data set Season

Month	State	By_season
3	Alaska	Winter
2	Alaska	Winter
3	Alaska	Winter
11	Alaska	Fall
4	Alaska	Spring
9	Alaska	Summer
5	Puerto Rico	Spring
11	Alaska	Fall
6	Alaska	Spring
5	Alaska	Spring
9	Alaska	Summer
4	Hawaii	Spring
1	California	Winter

Exercise# 3:

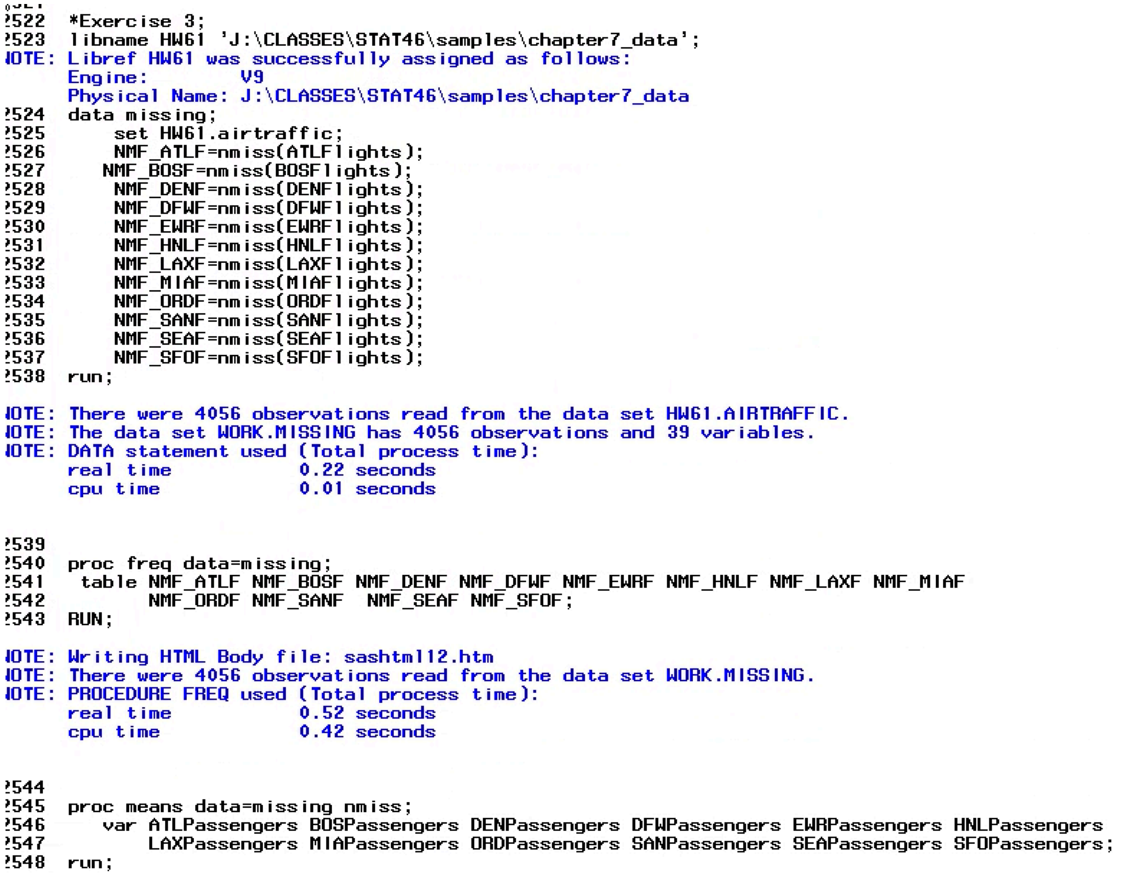
In SAS, as you may know by now, there are always various there are multiple ways to code for the same thing. In this exercise, I present two different ways to use for missing values: for the 1st question, NMISS is used as a function and for the second question, I use NMISS as an option with proc means.
One is more tedious than the other but they are both correct.

Program:

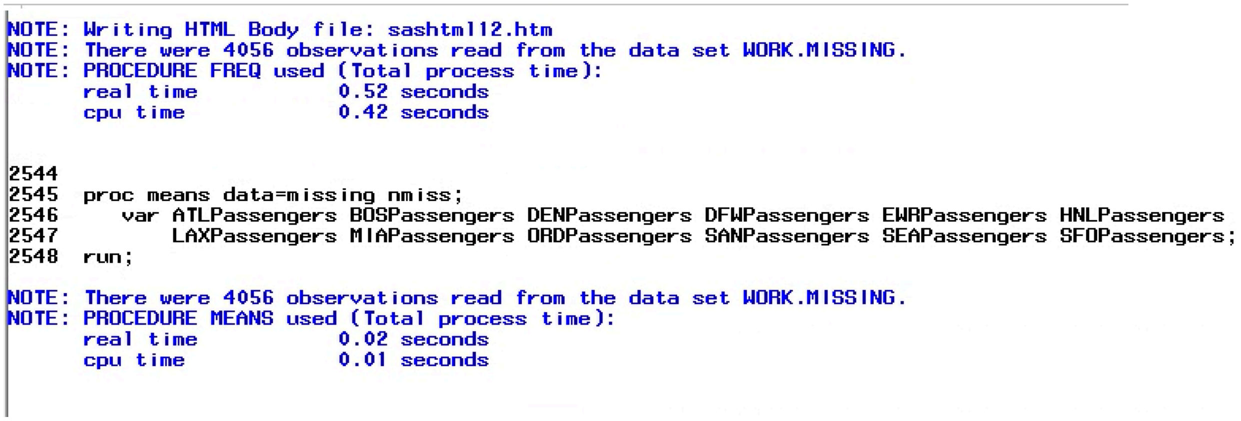


```
*Exercise 3;
libname HW61 'J:\CLASSES\STAT46\samples\chapter7_data';

*Method# 1;
data missing;
    set HW61.airtraffic;
    NMF_ATLF=nmiss(ATLFlights);
   NMF_BOSF=nmiss(BOSFlights);
    NMF_DENF=nmiss(DENFlights);
    NMF_DFWF=nmiss(DFWFlights);
    NMF_EWRF=nmiss(EWRFlights);
    NMF_HNLF=nmiss(HNLFlights);
    NMF_LAXF=nmiss(LAXFlights);
    NMF_MIAF=nmiss(MIAFlights);
    NMF_ORDF=nmiss(ORDFlights);
    NMF_SANF=nmiss(SANFlights);
    NMF_SEAF=nmiss(SEAFlights);
    NMF_SFOF=nmiss(SFOFlights);
run;

proc freq data=missing;
 table NMF_ATLF NMF_BOSF NMF_DENF NMF_DFWF NMF_EWRF NMF_HNLF NMF_LAXF NMF_MIAF
       NMF_ORDF NMF_SANF  NMF_SEAF NMF_SFOF;
RUN;

*Method# 2;

proc means data=missing nmiss;
   var ATLPassengers BOSPassengers DENPassengers DFWPassengers EWRPassengers HNLPassengers
       LAXPassengers MIAPassengers ORDPassengers SANPassengers SEAPassengers SFOPassengers;
run;
```

Log:

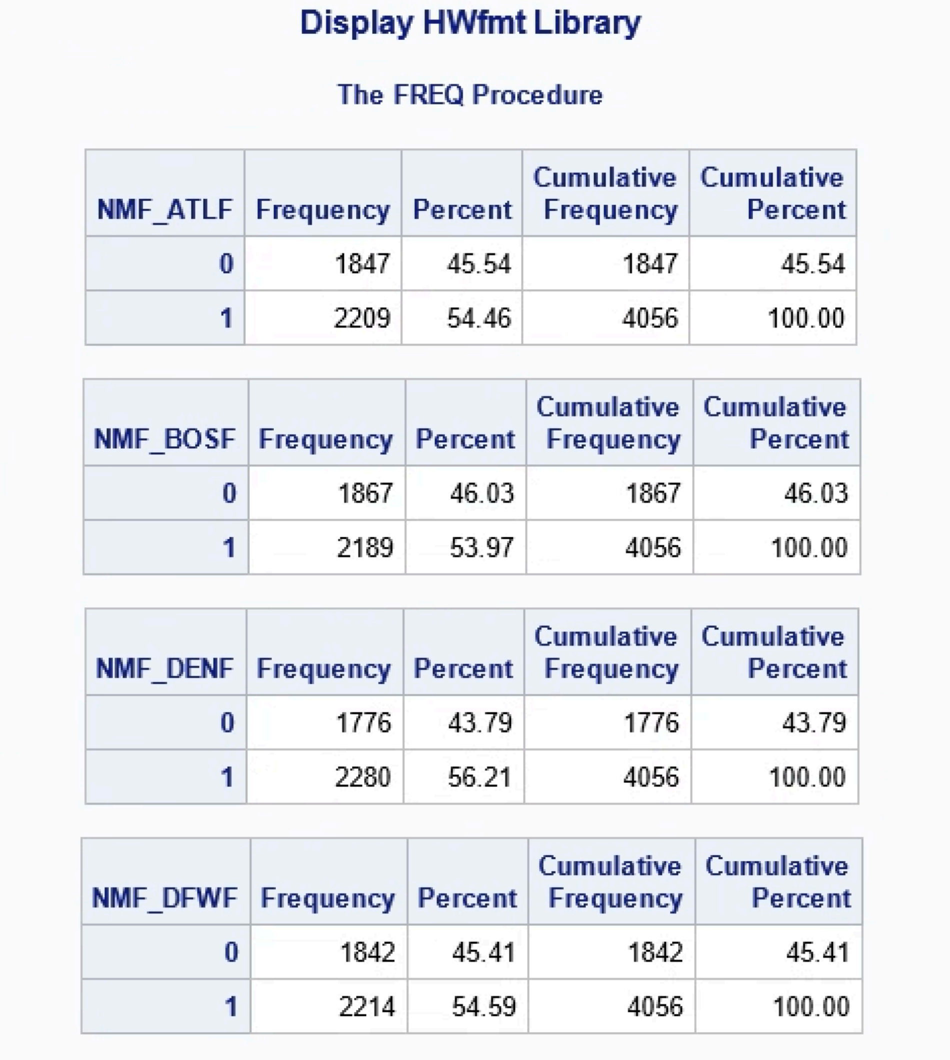


```
2522  *Exercise 3;
2523  libname HW61 'J:\CLASSES\STAT46\samples\chapter7_data';
NOTE: Libref HW61 was successfully assigned as follows:
      Engine:        V9
      Physical Name: J:\CLASSES\STAT46\samples\chapter7_data
2524  data missing;
2525      set HW61.airtraffic;
2526      NMF_ATLF=nmiss(ATLFlights);
2527     NMF_BOSF=nmiss(BOSFlights);
2528      NMF_DENF=nmiss(DENFlights);
2529      NMF_DFWF=nmiss(DFWFlights);
2530      NMF_EWRF=nmiss(EWRFlights);
2531      NMF_HNLF=nmiss(HNLFlights);
2532      NMF_LAXF=nmiss(LAXFlights);
2533      NMF_MIAF=nmiss(MIAFlights);
2534      NMF_ORDF=nmiss(ORDFlights);
2535      NMF_SANF=nmiss(SANFlights);
2536      NMF_SEAF=nmiss(SEAFlights);
2537      NMF_SFOF=nmiss(SFOFlights);
2538  run;

NOTE: There were 4056 observations read from the data set HW61.AIRTRAFFIC.
NOTE: The data set WORK.MISSING has 4056 observations and 39 variables.
NOTE: DATA statement used (Total process time):
      real time           0.22 seconds
      cpu time            0.01 seconds


2539
2540  proc freq data=missing;
2541   table NMF_ATLF NMF_BOSF NMF_DENF NMF_DFWF NMF_EWRF NMF_HNLF NMF_LAXF NMF_MIAF
2542         NMF_ORDF NMF_SANF  NMF_SEAF NMF_SFOF;
2543  RUN;

NOTE: Writing HTML Body file: sashtml12.htm
NOTE: There were 4056 observations read from the data set WORK.MISSING.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.52 seconds
      cpu time            0.42 seconds


2544
2545  proc means data=missing nmiss;
2546     var ATLPassengers BOSPassengers DENPassengers DFWPassengers EWRPassengers HNLPassengers
2547         LAXPassengers MIAPassengers ORDPassengers SANPassengers SEAPassengers SFOPassengers;
2548  run;
```



```
NOTE: Writing HTML Body file: sashtml12.htm
NOTE: There were 4056 observations read from the data set WORK.MISSING.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.52 seconds
      cpu time            0.42 seconds


2544
2545  proc means data=missing nmiss;
2546     var ATLPassengers BOSPassengers DENPassengers DFWPassengers EWRPassengers HNLPassengers
2547         LAXPassengers MIAPassengers ORDPassengers SANPassengers SEAPassengers SFOPassengers;
2548  run;

NOTE: There were 4056 observations read from the data set WORK.MISSING.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.02 seconds
      cpu time            0.01 seconds
```

Output Method# 1:



Display HWfmt Library

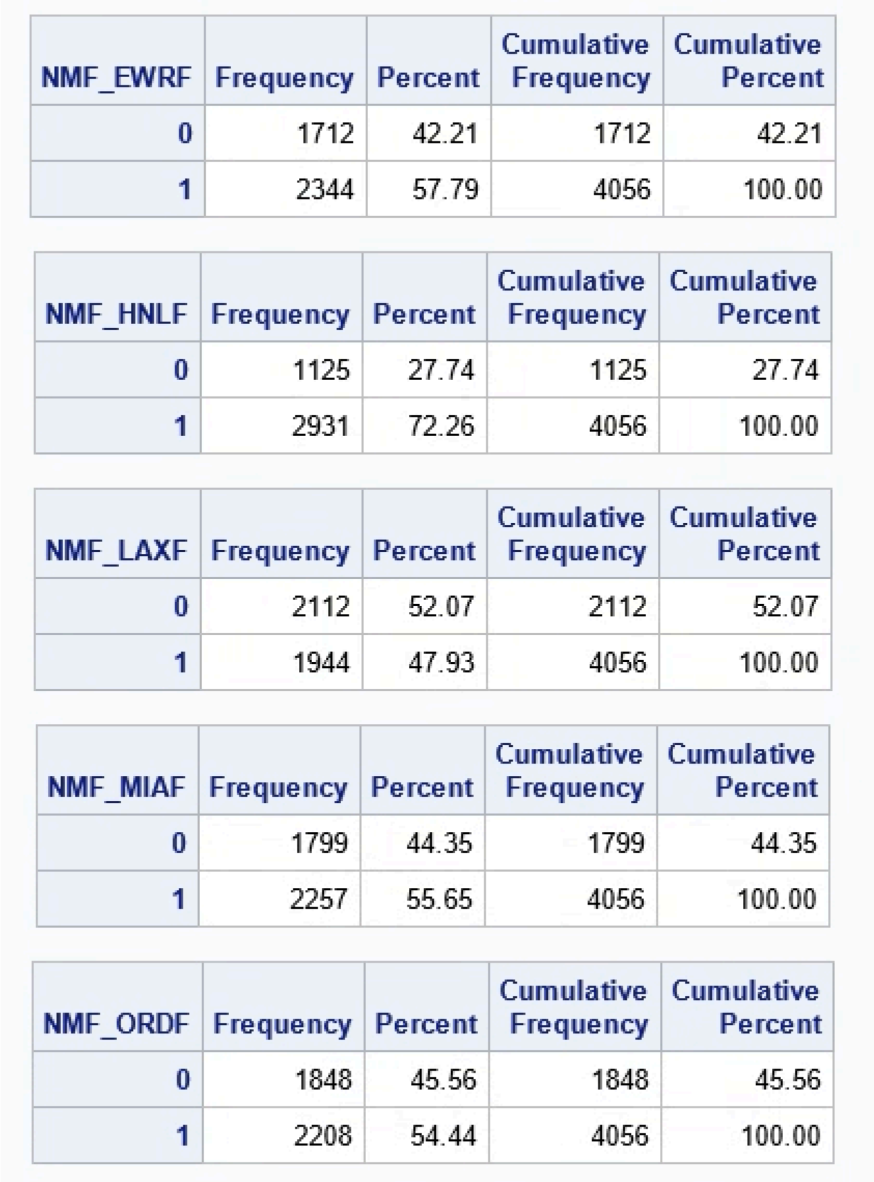
The FREQ Procedure

NMF_ATLF	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1847	45.54	1847	45.54
1	2209	54.46	4056	100.00

NMF_BOSF	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1867	46.03	1867	46.03
1	2189	53.97	4056	100.00

NMF_DENF	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1776	43.79	1776	43.79
1	2280	56.21	4056	100.00

NMF_DFWF	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1842	45.41	1842	45.41
1	2214	54.59	4056	100.00



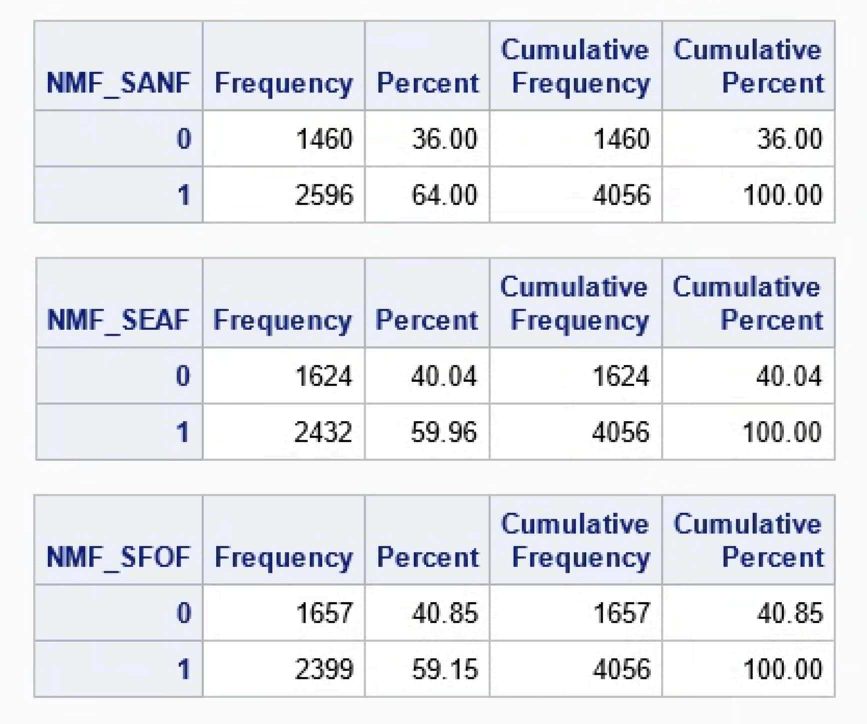
NMF_EWRF	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1712	42.21	1712	42.21
1	2344	57.79	4056	100.00

NMF_HNLF	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1125	27.74	1125	27.74
1	2931	72.26	4056	100.00

NMF_LAXF	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	2112	52.07	2112	52.07
1	1944	47.93	4056	100.00

NMF_MIAF	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1799	44.35	1799	44.35
1	2257	55.65	4056	100.00

NMF_ORDF	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1848	45.56	1848	45.56
1	2208	54.44	4056	100.00



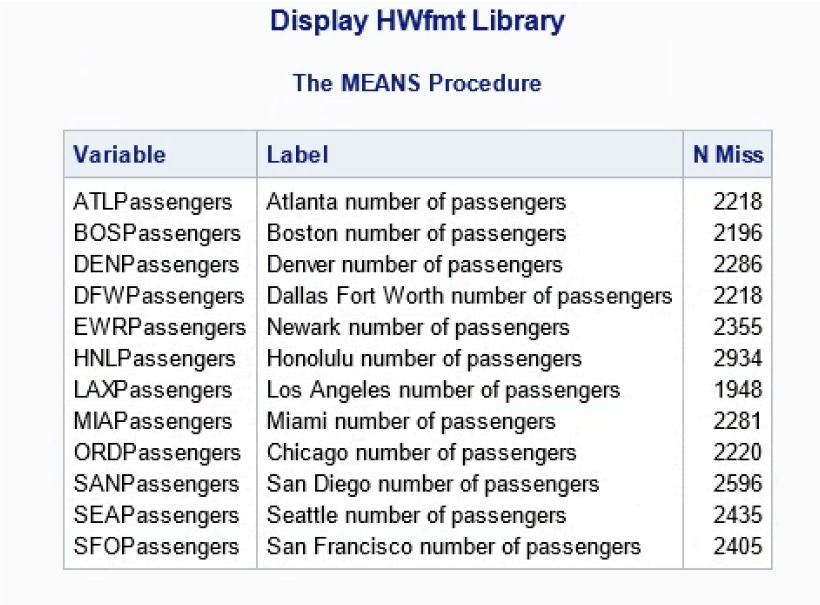
NMF_SANF	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1460	36.00	1460	36.00
1	2596	64.00	4056	100.00

NMF_SEAF	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1624	40.04	1624	40.04
1	2432	59.96	4056	100.00

NMF_SFOF	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1657	40.85	1657	40.85
1	2399	59.15	4056	100.00

Using **NMISS** as a function, we can use **proc freq**, where 1 represents the missing flight. So we can see by reviewing the frequency tables that Honolulu has the maximum missing values for the number of flights.

Output Method# 2:



Display HWfmt Library

The MEANS Procedure

Variable	Label	N Miss
ATLPassengers	Atlanta number of passengers	2218
BOSPassengers	Boston number of passengers	2196
DENPassengers	Denver number of passengers	2286
DFWPassengers	Dallas Fort Worth number of passengers	2218
EWRPassengers	Newark number of passengers	2355
HNLPassengers	Honolulu number of passengers	2934
LAXPassengers	Los Angeles number of passengers	1948
MIAPassengers	Miami number of passengers	2281
ORDPassengers	Chicago number of passengers	2220
SANPassengers	San Diego number of passengers	2596
SEAPassengers	Seattle number of passengers	2435
SFOPassengers	San Francisco number of passengers	2405

Using **proc means** with **NMISS** as an option, we have one table that shows all the missing values for the number of passengers, and we can see that Los Angeles is the winning city.