

```

*STEP 1 ;
/*****
*****/
/*1. Program Name:Vivek235_HW04_Program.sas
*/
/* Program Location: C:\Users\vigupta\OneDrive\Learning\DataScience\Statistics Texas A&M
University\657\Homework\Assignment04\Assignment\Vivek235_HW04_Program.sas */
/* Date Created: 2/1/17
*/
/* Author: Vivek Kumar Gupta
*/
/* Purpose: This assignment will primarily utilize, but is not limited to, techniques covered in the first
5 lectures.
*/
/*****
*****/

*STEP 1 ;
*1.Create the necessary library references for data sources and destination and file references for output
;
libname ncaa 'C:\Users\vigupta\OneDrive\Learning\DataScience\Statistics Texas A&M
University\657\Homework\Assignment04\SourceData' access=readonly;
filename pdfdev 'C:\Users\vigupta\OneDrive\Learning\DataScience\Statistics Texas A&M
University\657\Homework\Assignment04\Vivek235_HW04_Output.pdf';

/*STEP 2. Open a PDF destination to capture the output from the procedures that follow. Create bookmarks
and hide them by default.*/
ods pdf file= pdfdev bookmarklist=hide style=ocean;

/* STEP 3. Concatenate ncaam03 and ncaam04 */

data work.ncaacombed ;
    set ncaa.ncaam03 ncaa.ncaam04;
run;

title"Top Teams from 2003 and 2004 Men's NCAA Tournaments";
title2"Concatenated Data";
/*STEP 4. Use SQL to print the data portion of the new data set. */
proc sql ;
    select * from work.ncaacombed ;
quit;

```

```

/*STEP 5. Interleave ncaam03 and ncaam04*/

/*Interleaving requires sorting thus sort the input datasets into
work library by player and team */
proc sort data=ncaa.ncaam03
    out=work.ncaam03_sorted;
    by player team;
run;

proc sort data=ncaa.ncaam04
    out=work.ncaam04_sorted;
    by player team;
run;

proc sort data=ncaa.ncaam06 (rename=(School=team))
    out=work.ncaam06_sorted;
    by player team;
run;

/*Run the concatenation/interleaving dropping the not needed variables */
data work.ncaacombinedinterleaved;
    set work.ncaam03_sorted(drop=region)
        work.ncaam04_sorted(drop=f3);
by player team;
run;

/*Set the required title */
title"Top Teams from 2003 and 2004 Men's NCAA Tournaments";
title2"First 30 Records of Interleaved Data";

/* STEP 6.Use the Print procedure and a data set option to print the first 30 records of interleaved data
set.*/

proc print data=work.ncaacombinedinterleaved(obs=30);
run;

/*STEP 7.Use the Match Merge process to create a data set of only those who played in both 2003 and 2004
tournaments. Again exclude any variables that are not in both data sets.*/

data work.ncaa03and04 (drop = region f3);
    merge work.ncaam03_sorted(in=ncaa03)

```

```

        work.ncaam04_sorted(in=ncaa04);
    by player team;
    if ncaa03=1 and ncaa04=1;
    run;

/*Set the required title */
title"Players Who Played in Both 2003 and 2004 Tournaments";
title2;
title3"NOTE: PPG is from 2003";

footnote"PPG is from 2003 since the dataset ncaam03 is listed first in the merge step. Order of the dataset
is important here.";

/*STEP 8. Use SQL to print Player, Team, and PPG from the merged data. The SQL statement must sort the list
by
descending PPG. The title and footnote shown in the sample output are only place holders for the actual
title
and footnote that you will use in your solution. Since PPG is in both data sets, your title must specify
from which
year the PPG value was taken. The footnote must provide a very brief explanation of why this year was used
for
PPG.*/
proc sql ;
select  player as Player , team as Team , ppg as PPG
from work.ncaa03and04
order by ppg desc;
quit;

/*STEP 9. Match Merge ncaam03, ncaam04, and ncaam06 into a single data set. */
data work.ncaa03and04and06 (keep = team player ppg2003 ppg2004 ppg2006);
    length player $19;
    merge work.ncaam03_sorted(in=ncaa03  rename=(ppg=ppg2003))
          work.ncaam04_sorted(in=ncaa04  rename=(ppg=ppg2004))
          work.ncaam06_sorted(in=ncaa06  rename=(ppg=ppg2006));
    by player team;
    label ppg2003= '2003 PPG'
          ppg2004 = '2004 PPG'
          ppg2006 = '2006 PPG';
    run;

```

/\*STEP 10.Create a SQL procedure with multiple statements. The first statement must write to the log a list of the columns and their attributes from the data set created in the previous step. The second statement will print the data portion of the data set with the columns in the order shown. Use this statement to assign the column labels shown in the sample output. \*/

```
title"Three-year NCAA Tournament Statistics";
footnote;
  proc sql feedback;
  select player , team , ppg2003 ,ppg2004, ppg2006
  from work.ncaa03and04and06;
quit;
```

/\*STEP 11. Use a single PROC step to print the descriptor portion of all data sets in your work library\*/

```
title"Descriptor Portion of Data Sets in the Work Library";
footnote;

proc contents data=work._all_;
run;
```

/\*STEP 12 Housekeeping\*/

```
title;
footnote;
```

```
/*Close the ODS destination*/
ods pdf close;
```