```
1780  *STEP 1 ;
1781  /*************************************************************************************
1781! *********************************/
1782  /*1. Program Name:Vivek235_HW04_Program.sas
1782!                                    */
1783  /* Program Location: C:\Users\vigupta\OneDrive\Learning\DataScience\Statistics Texas A&M
1783! University\657\Homework\Assignment04\Assignment\Vivek235_HW04_Program.sas      */
1784  /* Date Created: 2/1/17
1784!                              */
1785  /* Author: Vivek Kumar Gupta
1785!                                 */
1786  /* Purpose: This assignment will primarily utilize, but is not limited to, techniques
1786! covered in the first 5 lectures.                       */
1787  /*************************************************************************************
1787! *********************************/
1788
1789  *STEP 1 ;
1790  *1.Create the necessary library references for data sources and destination and file
1790! references for output ;
1791  libname ncaa  'C:\Users\vigupta\OneDrive\Learning\DataScience\Statistics Texas A&M
1791! University\657\Homework\Assignment04\SourceData' access=readonly;
NOTE: Libref NCAA was successfully assigned as follows:
      Engine:        V9
      Physical Name: C:\Users\vigupta\OneDrive\Learning\DataScience\Statistics Texas A&M
      University\657\Homework\Assignment04\SourceData
1792  filename pdfdev 'C:\Users\vigupta\OneDrive\Learning\DataScience\Statistics Texas A&M
1792! University\657\Homework\Assignment04\Vivek235_HW04_Output.pdf';
1793
1794  /*STEP 2. Open a PDF destination to capture the output from the procedures that follow.
1794! Create bookmarks and hide them by default.*/
1795  ods pdf file= pdfdev  bookmarklist=hide style=ocean;
NOTE: Writing ODS PDF output to DISK destination "PDFDEV", printer "PDF".
1796
1797  /* STEP 3. Concatenate ncaam03 and ncaam04 */
1798
1799  data work.ncaacombined ;
1800      set ncaa.ncaam03 ncaa.ncaam04;
1801  run;

NOTE: There were 267 observations read from the data set NCAA.NCAAM03.
NOTE: There were 73 observations read from the data set NCAA.NCAAM04.
```

```
NOTE: The data set WORK.NCAACOMBINED has 340 observations and 10 variables.
NOTE: DATA statement used (Total process time):
      real time           0.02 seconds
      cpu time            0.01 seconds


1802
1803  title"Top Teams from 2003 and 2004 Men's NCAA Tournaments";
1804  title2"Concatenated Data";
1805  /*STEP 4. Use SQL to print the data portion of the new data set. */
1806  proc sql ;
1807   select * from work.ncaacombined ;
1808  quit;
NOTE: PROCEDURE SQL used (Total process time):
      real time           0.27 seconds
      cpu time            0.26 seconds


1809
1810  /*STEP 5. Interleave ncaam03 and ncaam04*/
1811
1812  /*Interleaving requires sorting thus sort the input datasets into
1813  work library by player and team */
1814  proc sort data=ncaa.ncaam03
1815      out=work.ncaam03_sorted;
1816      by player team;
1817  run;

NOTE: There were 267 observations read from the data set NCAA.NCAAM03.
NOTE: The data set WORK.NCAAM03_SORTED has 267 observations and 9 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.02 seconds
      cpu time            0.01 seconds


1818
1819  proc sort data=ncaa.ncaam04
1820      out=work.ncaam04_sorted;
1821      by player team;
1822  run;
```

```
NOTE: There were 73 observations read from the data set NCAA.NCAAM04.
NOTE: The data set WORK.NCAAM04_SORTED has 73 observations and 9 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.02 seconds
      cpu time            0.01 seconds


1823
1824  proc sort data=ncaa.ncaam06 (rename=(School=team))
1825      out=work.ncaam06_sorted;
1826      by player team;
1827  run;

NOTE: There were 193 observations read from the data set NCAA.NCAAM06.
NOTE: The data set WORK.NCAAM06_SORTED has 193 observations and 13 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


1828
1829  /*Run the concatenation/interleaving dropping the not needed variables */
1830  data work.ncaacombinedinterleaved;
1831      set work.ncaam03_sorted(drop=region)
1832          work.ncaam04_sorted(drop=f3);
1833  by player team;
1834  run;

NOTE: There were 267 observations read from the data set WORK.NCAAM03_SORTED.
NOTE: There were 73 observations read from the data set WORK.NCAAM04_SORTED.
NOTE: The data set WORK.NCAACOMBINEDINTERLEAVED has 340 observations and 8 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


1835
1836  /*Set the required title */
1837  title"Top Teams from 2003 and 2004 Men's NCAA Tournaments";
1838  title2"First 30 Records of Interleaved Data";
1839
```

```
1840  /* STEP 6.Use the Print procedure and a data set option to print the first 30 records of
1840! interleaved data set.*/
1841
1842  proc print data=work.ncaacombinedinterleaved(obs=30);
1843  run;

NOTE: There were 30 observations read from the data set WORK.NCAACOMBINEDINTERLEAVED.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.04 seconds
      cpu time            0.04 seconds


1844
1845  /*STEP 7.Use the Match Merge process to create a data set of only those who played in both
1845! 2003 and 2004
1846  tournaments. Again exclude any variables that are not in both data sets.*/
1847
1848  data work.ncaa03and04 (drop = region f3);
1849      merge work.ncaam03_sorted(in=ncaa03)
1850          work.ncaam04_sorted(in=ncaa04);
1851      by player team;
1852      if ncaa03=1 and ncaa04=1;
1853      run;

NOTE: There were 267 observations read from the data set WORK.NCAAM03_SORTED.
NOTE: There were 73 observations read from the data set WORK.NCAAM04_SORTED.
NOTE: The data set WORK.NCAA03AND04 has 29 observations and 8 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


1854
1855  /*Set the required title */
1856  title"Players Who Played in Both 2003 and 2004 Tournaments";
1857  title2;
1858  title3"NOTE: PPG is from 2003";
1859
1860  footnote"PPG is from 2003 since the dataset ncaam03 is listed first in the merge step. Order
1860!  of the dataset is important here.";
1861
```

```
1862  /*STEP 8. Use SQL to print Player, Team, and PPG from the merged data. The SQL statement
1862! must sort the list by
1863  descending PPG. The title and footnote shown in the sample output are only place holders for
1863!  the actual title
1864  and footnote that you will use in your solution. Since PPG is in both data sets, your title
1864! must specify from which
1865  year the PPG value was taken. The footnote must provide a very brief explanation of why this
1865!  year was used for
1866  PPG.*/
1867  proc sql ;
1868  select  player as Player , team as Team , ppg as PPG
1869  from work.ncaa03and04
1870  order by ppg desc;
1871  quit;
NOTE: PROCEDURE SQL used (Total process time):
      real time           0.03 seconds
      cpu time            0.03 seconds


1872
1873
1874  /*STEP 9. Match Merge ncaam03, ncaam04, and ncaam06 into a single data set. */
1875  data work.ncaa03and04and06 (keep = team player ppg2003 ppg2004 ppg2006);
1876      length player $19;
1877      merge work.ncaam03_sorted(in=ncaa03  rename=(ppg=ppg2003))
1878          work.ncaam04_sorted(in=ncaa04 rename=(ppg=ppg2004))
1879          work.ncaam06_sorted(in=ncaa06 rename=(ppg=ppg2006));
1880      by player team;
1881      label ppg2003= '2003 PPG'
1882      ppg2004 ='2004 PPG'
1883      ppg2006 ='2006 PPG';
1884      run;

WARNING: Multiple lengths were specified for the BY variable player by input data sets and
         LENGTH, FORMAT, INFORMAT, or ATTRIB statements. This might cause unexpected results.
NOTE: There were 267 observations read from the data set WORK.NCAAM03_SORTED.
NOTE: There were 73 observations read from the data set WORK.NCAAM04_SORTED.
NOTE: There were 193 observations read from the data set WORK.NCAAM06_SORTED.
NOTE: The data set WORK.NCAA03AND04AND06 has 491 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           0.02 seconds
```

```
      cpu time            0.01 seconds


1885
1886  /*STEP 10.Create a SQL procedure with multiple statements. The first statement must write to
1886!  the log a list of the columns
1887  and their attributes from the data set created in the previous step. The second statement
1887! will print the data
1888  portion of the data set with the columns in the order shown. Use this statement to assign
1888! the column labels
1889  shown in the sample output. */
1890
1891  title"Three-year NCAA Tournament Statistics";
1892  footnote;
1893       proc sql feedback;
1894       select player , team , ppg2003 ,ppg2004, ppg2006
1895       from work.ncaa03and04and06;
NOTE: Statement transforms to:

        select NCAA03AND04AND06.player, NCAA03AND04AND06.Team, NCAA03AND04AND06.ppg2003,
NCAA03AND04AND06.ppg2004, NCAA03AND04AND06.ppg2006
          from WORK.NCAA03AND04AND06;

1896       quit;
NOTE: PROCEDURE SQL used (Total process time):
      real time           0.19 seconds
      cpu time            0.18 seconds


1897
1898  /*STEP 11. Use a single PROC step to print the descriptor portion of all data sets in your
1898! work library*/
1899
1900       title"Descriptor Portion of Data Sets in the Work Library";
1901       footnote;
1902
1903       proc contents data=work._all_;
1904       run;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.13 seconds
```

```
      cpu time            0.12 seconds


1905
1906  /*STEP 12 Housekeeping*/
1907      title;
1908      footnote;
1909
1910  /*Close the ODS destination*/
1911  ods pdf close;
NOTE: ODS PDF printed 37 pages to C:\Users\vigupta\OneDrive\Learning\DataScience\Statistics Texas
      A&M University\657\Homework\Assignment04\Vivek235_HW04_Output.pdf.
```