```
1    *STEP 0 ;
2    /*1. Program Name:Vivek235_HW13_Program.sas.
2  !                                      */
3    /* Program Location: C:\Users\vigupta\OneDrive\Learning\DataScience\Statistics Texas A&M
3  ! University\657\Homework\Assignment12\Vivek235_HW13_Program.sas     */
4    /* Date Created: 4/13/17
4  !                                */
5    /* Author: Vivek Kumar Gupta
5  !                                */
6    /* Purpose:This assignment focuses on the techniques taught in the last few lectures of the
6  ! semester. You
7    should have all of the information you need by the end of Lecture 22. You will be using the
8    ncaam06 and ncaam04 datasets that were used in previous assignments.;           */
9
/*************************************************************************************************
9  ! *******************************/
10
11   /*Set up appropriate libnames need for the program*/
12   libname srcdata 'C:\Users\vigupta\OneDrive\Learning\DataScience\Statistics Texas A&M
12 ! University\657\Homework\Assignment04\SourceData' access=readonly;
NOTE: Libref SRCDATA was successfully assigned as follows:
      Engine:        V9
      Physical Name: C:\Users\vigupta\OneDrive\Learning\DataScience\Statistics Texas A&M
      University\657\Homework\Assignment04\SourceData
13   filename pdfdev 'C:\Users\vigupta\OneDrive\Learning\DataScience\Statistics Texas A&M
13 ! University\657\Homework\Assignment13\Vivek235_HW13_Output.pdf';
14
15   /*Setp up system level options*/
16   option msglevel = i nodate nonumber;
17
18   /*Setp up the ODS for the output*/
19   ods pdf file=pdfdev  bookmarkgen=no ;
NOTE: Writing ODS PDF output to DISK destination "PDFDEV", printer "PDF".
20
21   /*STEP 1. Use PROC SQL to create a table with columns seed, school, region, player, ppg, and
21 ! rpg from ncaam06 with
22   only schools that have 5 or more players listed in the dataset.*/
23   proc sql ;
24   /*STEP 1a. SQL commands to create the dataset*/
25   /*STEP 1b. The Start value of your format will come from the Team column.*/
26   /*STEP 1c. Use seed_ as the label in your format.*/
27   /*STEP 1d. Give the new format a name of your choosing*/
28   create table TeamSeed as
29   select distinct team as Start , put(seed_, 2.) as Label, "$TeamSeed" as FmtName from
29 ! srcdata.ncaam04
30   order by label;
NOTE: Table WORK.TEAMSEED created, with 15 rows and 3 columns.

31
32   /*STEP 1e. After you have created the new dataset, use SQL to insert a row at the end. This
32 ! row will
33   have values of start='other' and label='NA' along with the name of your format*/
34
35   insert into TeamSeed(Start, Label,FmtName)
36   values("other","NA","$TeamSeed");
NOTE: 1 row was inserted into WORK.TEAMSEED.
```

```
37   quit;
NOTE: PROCEDURE SQL used (Total process time):
      real time           0.07 seconds
      cpu time            0.06 seconds


38
39   /*STEP 1f. Use the format procedure to create from this dataset a user defined format in the
39 ! work
40   library.*/
41   proc format library=work
42        cntlin=TeamSeed ;
NOTE: Format $TEAMSEED has been output.
43        select FmtName;
44   run;

NOTE: PROCEDURE FORMAT used (Total process time):
      real time           0.04 seconds
      cpu time            0.01 seconds

NOTE: No formats found that match SELECT statement.
NOTE: There were 16 observations read from the data set WORK.TEAMSEED.

45
46   /*STEP 1g. Use the format procedure to write the contents of the new format out to your
46 ! output
47   document.*/
48   proc format library=work fmtlib;
49       select $TeamSeed;
50   run;

NOTE: PROCEDURE FORMAT used (Total process time):
      real time           0.44 seconds
      cpu time            0.20 seconds

NOTE: Writing HTML Body file: sashtml.htm
NOTE: Non-portable document will be produced. The current settings of FORMCHAR use nonstandard
      line-drawing characters and the resulting output file will not render correctly unless all
      readers of the document have the SAS Monospace font installed. To make your document
      portable, issue the following command:
      OPTIONS FORMCHAR="|----|+|---+=|-/\<>*";


51
52   /*STEP 2. Create a picture format that can be applied to PPG column*/
53   proc format library=work  ;
54   picture ppg
55   15-high = '(09.9)' (prefix='High(')
56   7.7 - < 15 = '(9.9)' (prefix='Medium(')
57   low - <7.7 ='(9.9)' (prefix='Low(');
NOTE: Format PPG has been output.
58
59   run;

NOTE: PROCEDURE FORMAT used (Total process time):
      real time           0.02 seconds
      cpu time            0.01 seconds
```

```
60
61   /*STEP 2. Use a SAS procedure to place a copy of the ncaam06 dataset in the WORK library.
All
61 !  future
62   references to ncaam06 will be to the WORK copy.*/
63
64   title "Descriptor Portion of ncaam06";
65
66   proc datasets library=srcdata nolist mtype=data;
67   copy out = work noclone;
68   select ncaam06;
69   run;

NOTE: Copying SRCDATA.NCAAM06 to WORK.NCAAM06 (memtype=DATA).
NOTE: System Options for BUFSIZE and REUSE were used at user's request.
NOTE: Libname and/or system options for compress, pointobs, data representation and encoding
      attributes were used at user's request.
INFO: Engine's block-read method is in use.
INFO: Engine's block-write  method is in use.
NOTE: There were 193 observations read from the data set SRCDATA.NCAAM06.
NOTE: The data set WORK.NCAAM06 has 193 observations and 13 variables.
70
71

NOTE: PROCEDURE DATASETS used (Total process time):
      real time           0.06 seconds
      cpu time            0.03 seconds


72   proc sql;
73
74   /*STEP 4a Create a character column labeled PPG Rating. This column needs to be wide enough
74 ! to
75   accept all values of the picture format label created above.*/
76
77   /*STEP 4b Create a character column with a length of 2 labeled 2004 Seeding.*/
78   /*STEP 4c. Change the length and format of the School column so that it is wide enough to
78 ! fully
79   display Wisconsin Milwaukee.*/
80
81   alter table work.ncaam06
82   add ppgchar char(20) label "PPG Rating"
83   add seed_04 char(2) label "2004 Seeding"
84   modify School char(40);
NOTE: Table WORK.NCAAM06 has been modified, with 15 columns.
85
86   /*STEP 4d Populate the PPG Rating column using the put function to create values by applying
86 ! the
87   picture format to the PPG column.*/
88   /*STE 4e. Correct the following school names*/
89   /*STEP 4f. Populate the 2004 Seeding column using the put function to create values by
89 ! applying
90   the format from step 1 to the school column.*/
91
92   update work.ncaam06
```

```
93   set ppgchar=put(ppg, ppg.),
94   School =
95       case(School)
96       when 'Indania' then 'Indiana'
97       when 'Boston Coll' then 'Boston College'
98       when 'George mason' then 'George Mason'
99       when 'Oral Robt-16' then 'Oral Roberts'
100      when 'Wisc. Milwaukee' then 'Wisconsin Milwaukee'
101      else School
102  end;
NOTE: 193 rows were updated in WORK.NCAAM06.

103
104  update work.ncaam06
105  set seed_04 = put(school,$TeamSeed.);
NOTE: 193 rows were updated in WORK.NCAAM06.

106
107  /*STEP 5. Create a composite index for ncaam06 using the variables player, school, and
region
107!  in that
108  order. */
109  create index comp_idx on work.ncaam06(player, school, region);
NOTE: Composite index comp_idx has been defined.
110
111  quit;
NOTE: PROCEDURE SQL used (Total process time):
      real time           0.15 seconds
      cpu time            0.06 seconds


112
113  /*STEP 5 Print the descriptor portion of the ncaam06 data set after the index has been
113! created.*/
114  proc datasets library=work nolist ;
115  contents data= ncaam06;
116  quit;

NOTE: PROCEDURE DATASETS used (Total process time):
      real time           0.07 seconds
      cpu time            0.04 seconds


117
118  /*STEP6.Question and answers */
119  title '6a. IDXWHERE on Player';
120
121  /*Yes the index will be used here because the composite index comp_idx
122  since we are forcing SAS to use the index regardless of whether sequential scan of the table
122! will be optimal or not
123  with the option idxwhere=yes.
124
125  The index will be used since comp_idx as player as the primary key*/
126
127  proc print data = work.ncaam06 (idxwhere=yes) label;
128  var Player School Region Seed ppgchar seed_04;
129  where player in ('Steve Burtt', 'Jared Dudley', 'Stanley Burrell');
```

```
INFO: Data set option (IDXWHERE=YES)forced an index to be used rather than a sequential pass for
where-clause processing.
INFO: Index comp_idx selected for WHERE clause optimization.
130  run;

NOTE: There were 3 observations read from the data set WORK.NCAAM06.
      WHERE player in ('Jared Dudley', 'Stanley Burrell', 'Steve Burtt');
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.05 seconds
      cpu time            0.03 seconds


131
132
133  title '6b. IDXWHERE on School';
134
135  /*This step produces an  error since we are forcing SAS to use an index but the only index
135! set on the table does not have school as
136  the primary key variable. By the rules of index usage, this is not allowed.
137  */
138
139  proc print data = work.ncaam06 (idxwhere=yes) label;
140  var Player School Region Seed ppgchar seed_04;
141  where school='Texas';
ERROR: IDXWHERE=YES but no index exists for optimizing a WHERE clause.
142  run;

NOTE: The SAS System stopped processing this step because of errors.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds

143
144
145  title '6c. IDXWHERE on School';
146
147  /* This step produces an error because we are fociing SAS to use an index that is not
147! applicable to the query.
148  */


149  proc print data = work.ncaam06 (idxname=comp_idx) label;
150  var Player School Region Seed ppgchar seed_04;
151  where school='Texas';
ERROR: IDXNAME=comp_idx, but first or only index variable does not match any optimizable WHERE
       clause condition.
152  run;

NOTE: The SAS System stopped processing this step because of errors.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds

153
154  title '6d. IDXWHERE on Player or School';
155
156  /* This is a case of compound optimization that does not occur
```

```
157  since both the where clause conditions are not connected by an AND operator . All other
157! requirements of compoud
158  optimization are satisifed however.
159  */
160


161  proc print data = work.ncaam06 (idxwhere=yes) label;
162  var Player School Region Seed ppgchar seed_04;
163  where player in ('Steve Burtt', 'Jared Dudley', 'Stanley Burrell') or school='Indiana';
164  run;

NOTE: There were 9 observations read from the data set WORK.NCAAM06.
      WHERE player in ('Jared Dudley', 'Stanley Burrell', 'Steve Burtt') or (school='Indiana');
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.05 seconds
      cpu time            0.01 seconds


165
166
167  title '6e. IDXWHERE on Player and School';
168
169  /* This is a case of compound optimization of the index.
170  Both the where clause conditions are connected by an AND operator . Keys used are the first
2
171  keys of the composite index and substr function begins at the beinning
172  */
173
174  proc print data = work.ncaam06 (idxwhere=yes) label;
175  var Player School Region Seed ppgchar seed_04;
176  where substr(player,1,1)='S' and
177  school in ('Duke', 'Oral Roberts', 'Iona', 'Boston College',
178  'Gonzaga');
INFO: Data set option (IDXWHERE=YES)forced an index to be used rather than a sequential pass for
where-clause processing.
INFO: Index comp_idx selected for WHERE clause optimization.
179  run;

NOTE: There were 5 observations read from the data set WORK.NCAAM06.
      WHERE (SUBSTR(player, 1, 1)='S') and school in ('Boston College', 'Duke', 'Gonzaga',
      'Iona', 'Oral Roberts');
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.06 seconds
      cpu time            0.04 seconds


180
181
182  /*Close the device*/
183  ods pdf close;
NOTE: ODS PDF printed 6 pages to C:\Users\vigupta\OneDrive\Learning\DataScience\Statistics Texas
      A&M University\657\Homework\Assignment13\Vivek235_HW13_Output.pdf.
184
185  /*House cleaning*/
186  title;
187  footnote;
```

```
188  option date number msglevel=n;
```