```
NOTE: This session is executing on the X64_10PRO  platform.



NOTE: Updated analytical products:

      SAS/STAT 14.3
      SAS/ETS 14.3
      SAS/OR 14.3
      SAS/IML 14.3
      SAS/QC 14.3

NOTE: Additional host information:

 X64_10PRO WIN 10.0.17763  Workstation

NOTE: SAS initialization used:
      real time           3.83 seconds
      cpu time            1.34 seconds

WARNING: One or more libraries specified in the concatenated library
SASHELP
WARNING: do not exist.  These libraries were removed from the
concatenation.
1
/*****************************************************************/
2    /* Program Name: Homework 5
*/
3    /* Program Location: C:\Users\dsingh\Dropbox\Stat_604\Homework\HW5
*/
4    /* Date Created: 6/8/2019
*/
5    /* Author: Dhruv Singh
*/
6    /* Purpose: Oklahoma Schools data cleaning
*/
7
8    libname ok_sch
'C:\Users\dsingh\Dropbox\Tamu\Stat_604\Homework\hwdata' access =
readonly;
NOTE: Libref OK_SCH was successfully assigned as follows:
      Engine:        V9
      Physical Name:
C:\Users\dsingh\Dropbox\Tamu\Stat_604\Homework\hwdata
9    libname ok_cln 'C:\Users\dsingh\Dropbox\Tamu\Stat_604\Homework\HW5'
access = readonly;
NOTE: Libref OK_CLN was successfully assigned as follows:
      Engine:        V9
      Physical Name: C:\Users\dsingh\Dropbox\Tamu\Stat_604\Homework\HW5
10   filename report
'C:\Users\dsingh\Dropbox\Tamu\Stat_604\Homework\HW5\DSingh_HW05_output.pd
f';
11
12
13   /* Step 1: reading in dataset */
14
15   /* Step 2a: Replacing county with blanks */
16   data schools;
17       set ok_sch.ok_schools;
18       newvar = tranwrd(county, 'COUNTY', '');
```

```
19   run;

NOTE: There were 1785 observations read from the data set
OK_SCH.OK_SCHOOLS.
NOTE: The data set WORK.SCHOOLS has 1785 observations and 9 variables.
NOTE: DATA statement used (Total process time):
      real time           0.07 seconds
      cpu time            0.06 seconds


20
21   * dropping original county var;
22   data schools2;
23       set schools;
24       drop county;
25   run;

NOTE: There were 1785 observations read from the data set WORK.SCHOOLS.
NOTE: The data set WORK.SCHOOLS2 has 1785 observations and 8 variables.
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      cpu time            0.01 seconds


26
27   * renaming new couunty var;
28   data schools2;
29       set schools2 (rename = (newvar=County));
30   run;

NOTE: There were 1785 observations read from the data set WORK.SCHOOLS2.
NOTE: The data set WORK.SCHOOLS2 has 1785 observations and 8 variables.
NOTE: DATA statement used (Total process time):
      real time           0.04 seconds
      cpu time            0.01 seconds


31
32   * the following temp data creates new vars for numeric grade vars;
33   data schools3;
34       set schools2;
35       newvar = 0;
36       select;
37           when (grade8 not in ('n/a', '*'))
38           newvar=input(grade8, comma6.);
39           otherwise newvar=.;
40       end;
41
42       newvar1 = 0;
43       select;
44           when (grade9 not in ('n/a', '*'))
45           newvar1=input(grade9, comma6.);
46           otherwise newvar1=.;
47       end;
48
49       newvar2 = 0;
50       select;
51           when (grade10 not in ('n/a', '*'))
52           newvar2=input(grade10, comma6.);
53           otherwise newvar2=.;
```

```
54       end;
55
56       newvar3 = 0;
57       select;
58           when (grade11 not in ('n/a', '*'))
59           newvar3=input(grade11, comma6.);
60           otherwise newvar3=.;
61       end;
62
63       newvar4 = 0;
64       select;
65           when (grade12 not in ('n/a', '*'))
66           newvar4=input(grade12, comma6.);
67           otherwise newvar4=.;
68       end;
69
70   run;

NOTE: There were 1785 observations read from the data set WORK.SCHOOLS2.
NOTE: The data set WORK.SCHOOLS3 has 1785 observations and 13 variables.
NOTE: DATA statement used (Total process time):
      real time           0.08 seconds
      cpu time            0.04 seconds


71
72
73   * here the following temp data step renames created vars, and drop
old char types;
74   data schools4;
75       set schools3;
76       drop grade8 grade9 grade10 grade11 grade12;
77   run;

NOTE: There were 1785 observations read from the data set WORK.SCHOOLS3.
NOTE: The data set WORK.SCHOOLS4 has 1785 observations and 8 variables.
NOTE: DATA statement used (Total process time):
      real time           0.06 seconds
      cpu time            0.03 seconds


78
79   data schools4;
80       set schools4 (rename=(newvar=Grade8 newvar1=Grade9
newvar2=Grade10 newvar3=Grade11
80 ! newvar4=Grade12));
81   run;

NOTE: There were 1785 observations read from the data set WORK.SCHOOLS4.
NOTE: The data set WORK.SCHOOLS4 has 1785 observations and 8 variables.
NOTE: DATA statement used (Total process time):
      real time           0.04 seconds
      cpu time            0.01 seconds


82
83   * search and replace in string variables ;
84   data schools5;
85       set schools4;
86
```

```
87       /*Step 2c: renaming city variables to correct for mispellings */
88       select;
89           when (city = 'CHUOTEAU')
90           city=tranwrd(city, 'CHUOTEAU', 'CHOUTEAU');
91
92           when (city = 'OKC')
93           city=tranwrd(city, 'OKC', 'OKLAHOMA CITY');
94
95           when (city='JENKS')
96           city=tranwrd(city, 'JENKS', 'TULSA');
97
98           when (city='MUSKOGE')
99           city=tranwrd(city, 'MUSKOGE', 'MUSKOGEE');
100
101          when (city='RUSHSPRINGS')
102          city=tranwrd(city, 'RUSHSPRINGS', 'RUSH SPRINGS');
103
104          when (city='SEMIONOLE')
105          city=tranwrd(city, 'SEMIONOLE', 'SEMINOLE');
106
107          when (city='SO. COFFEEVILLE')
108          city=tranwrd(city, 'SO. COFFEEVILLE', 'SOUTH COFFEEVILLE');
109
110          when (city='WOOWARD')
111          city=tranwrd(city, 'WOOWARD', 'WOODWARD');
112
113          otherwise;
114      end;
115
116      /* Step 2d: renaming one county to combine it with another */
117      select;
118          when (county='ALFALFA')
119          county=tranwrd(county, 'ALFALFA', 'CHEROKEE');
120
121          otherwise;
122      end;
123  run;

NOTE: There were 1785 observations read from the data set WORK.SCHOOLS4.
NOTE: The data set WORK.SCHOOLS5 has 1785 observations and 8 variables.
NOTE: DATA statement used (Total process time):
      real time           0.07 seconds
      cpu time            0.06 seconds


124
125
126  /*Step 3a: Sorting data set for by group processing*/
127  proc sort data=schools5 out=summary;
128      by City;
129  run;

NOTE: There were 1785 observations read from the data set WORK.SCHOOLS5.
NOTE: The data set WORK.SUMMARY has 1785 observations and 8 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.04 seconds
      cpu time            0.01 seconds


130
```

```
131
132  /*Summarize Class Size by City*/
133  data summary (keep = City Grade8Sum Grade9Sum Grade10Sum Grade11Sum
Grade12Sum);
134       set summary;
135       by City;
136       if First.City then Grade8Sum=0;
137       Grade8Sum+Grade8;
138
139       if First.City then Grade9Sum=0;
140       Grade9Sum+Grade9;
141
142       if First.City then Grade10Sum=0;
143       Grade10Sum+Grade10;
144
145       if First.City then Grade11Sum=0;
146       Grade11Sum+Grade11;
147
148       if First.City then Grade12Sum=0;
149       Grade12Sum+Grade12;
150       *step 3b: removing redundant observations;
151       if Last.City;
152  run;

NOTE: There were 1785 observations read from the data set WORK.SUMMARY.
NOTE: The data set WORK.SUMMARY has 435 observations and 6 variables.
NOTE: DATA statement used (Total process time):
      real time           0.06 seconds
      cpu time            0.06 seconds


153
154
155  data summary2;
156       set summary;
157       *step 3c: labelling summary variables;
158       label Grade8Sum = 'Eighth Graders'
159           Grade9Sum = 'Ninth Graders'
160           Grade10Sum = 'Tenth Graders'
161           Grade11Sum = 'Eleventh Graders'
162           Grade12Sum = 'Twelfth Graders';
163
164       * step 3d: creating new vars for current and projected
enrollment;
165       CurrentEnrollment = Sum(Grade9Sum, Grade10Sum, Grade11Sum,
Grade12Sum);
166       ProjectedEnrollment = Sum(Grade8Sum, Grade9Sum, Grade10Sum,
Grade11Sum);
167
168       PercentChange = ((ProjectedEnrollment -
CurrentEnrollment)/CurrentEnrollment)*100;
169  run;

NOTE: Division by zero detected at line 168 column 63.
City=ALBION Grade8Sum=9 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0
Grade12Sum=0 CurrentEnrollment=0
ProjectedEnrollment=9 PercentChange=. _ERROR_=1 _N_=6
NOTE: Division by zero detected at line 168 column 63.
City=AVANT Grade8Sum=6 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0 Grade12Sum=0
CurrentEnrollment=0
```

```
ProjectedEnrollment=6 PercentChange=. _ERROR_=1 _N_=22
NOTE: Division by zero detected at line 168 column 63.
City=BOWRING Grade8Sum=6 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0
Grade12Sum=0 CurrentEnrollment=0
ProjectedEnrollment=6 PercentChange=. _ERROR_=1 _N_=43
NOTE: Division by zero detected at line 168 column 63.
City=BOYNTON Grade8Sum=3 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0
Grade12Sum=0 CurrentEnrollment=0
ProjectedEnrollment=3 PercentChange=. _ERROR_=1 _N_=44
NOTE: Division by zero detected at line 168 column 63.
City=BRAMAN Grade8Sum=5 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0
Grade12Sum=0 CurrentEnrollment=0
ProjectedEnrollment=5 PercentChange=. _ERROR_=1 _N_=46
NOTE: Division by zero detected at line 168 column 63.
City=BYARS Grade8Sum=5 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0 Grade12Sum=0
CurrentEnrollment=0
ProjectedEnrollment=5 PercentChange=. _ERROR_=1 _N_=55
NOTE: Division by zero detected at line 168 column 63.
City=CLEO SPRINGS Grade8Sum=11 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0
Grade12Sum=0
CurrentEnrollment=0 ProjectedEnrollment=11 PercentChange=. _ERROR_=1
_N_=82
NOTE: Division by zero detected at line 168 column 63.
City=DEER CREEK Grade8Sum=13 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0
Grade12Sum=0 CurrentEnrollment=0
ProjectedEnrollment=13 PercentChange=. _ERROR_=1 _N_=107
NOTE: Division by zero detected at line 168 column 63.
City=FANSHAWE Grade8Sum=6 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0
Grade12Sum=0 CurrentEnrollment=0
ProjectedEnrollment=6 PercentChange=. _ERROR_=1 _N_=134
NOTE: Division by zero detected at line 168 column 63.
City=FORT SILL Grade8Sum=0 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0
Grade12Sum=0 CurrentEnrollment=0
ProjectedEnrollment=0 PercentChange=. _ERROR_=1 _N_=141
NOTE: Division by zero detected at line 168 column 63.
City=GARVIN Grade8Sum=16 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0
Grade12Sum=0 CurrentEnrollment=0
ProjectedEnrollment=16 PercentChange=. _ERROR_=1 _N_=151
NOTE: Division by zero detected at line 168 column 63.
City=GRANT Grade8Sum=17 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0
Grade12Sum=0 CurrentEnrollment=0
ProjectedEnrollment=17 PercentChange=. _ERROR_=1 _N_=161
NOTE: Division by zero detected at line 168 column 63.
City=HAYWOOD Grade8Sum=17 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0
Grade12Sum=0 CurrentEnrollment=0
ProjectedEnrollment=17 PercentChange=. _ERROR_=1 _N_=173
NOTE: Division by zero detected at line 168 column 63.
City=HITCHITA Grade8Sum=13 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0
Grade12Sum=0 CurrentEnrollment=0
ProjectedEnrollment=13 PercentChange=. _ERROR_=1 _N_=180
NOTE: Division by zero detected at line 168 column 63.
City=HODGEN Grade8Sum=19 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0
Grade12Sum=0 CurrentEnrollment=0
ProjectedEnrollment=19 PercentChange=. _ERROR_=1 _N_=182
NOTE: Division by zero detected at line 168 column 63.
City=JENNINGS Grade8Sum=24 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0
Grade12Sum=0 CurrentEnrollment=0
ProjectedEnrollment=24 PercentChange=. _ERROR_=1 _N_=196
NOTE: Division by zero detected at line 168 column 63.
```

```
City=JET Grade8Sum=17 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0 Grade12Sum=0
CurrentEnrollment=0
ProjectedEnrollment=17 PercentChange=. _ERROR_=1 _N_=197
NOTE: Division by zero detected at line 168 column 63.
City=KAW CITY Grade8Sum=22 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0
Grade12Sum=0 CurrentEnrollment=0
ProjectedEnrollment=22 PercentChange=. _ERROR_=1 _N_=200
NOTE: Division by zero detected at line 168 column 63.
City=KREBS Grade8Sum=31 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0
Grade12Sum=0 CurrentEnrollment=0
ProjectedEnrollment=31 PercentChange=. _ERROR_=1 _N_=211
NOTE: Division by zero detected at line 168 column 63.
WARNING: Limit set by ERRORS= option reached.  Further errors of this
type will not be printed.
City=LANE Grade8Sum=15 Grade9Sum=0 Grade10Sum=0 Grade11Sum=0 Grade12Sum=0
CurrentEnrollment=0
ProjectedEnrollment=15 PercentChange=. _ERROR_=1 _N_=215
NOTE: Mathematical operations could not be performed at the following
places. The results of the
      operations have been set to missing values.
      Each place is given by: (Number of times) at (Line):(Column).
      48 at 168:63
NOTE: There were 435 observations read from the data set WORK.SUMMARY.
NOTE: The data set WORK.SUMMARY2 has 435 observations and 9 variables.
NOTE: DATA statement used (Total process time):
      real time           0.10 seconds
      cpu time            0.06 seconds


170
171  /* Step 4: Creating output delivery system to pdf */
172  ods pdf file = report bookmarkgen = no;
NOTE: Writing ODS PDF output to DISK destination "REPORT", printer "PDF".
173
174  /* Step 5: Printing relevant data with labels created */
175  proc print data = schools5;
NOTE: Writing HTML Body file: sashtml.htm
176  run;

NOTE: There were 1785 observations read from the data set WORK.SCHOOLS5.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           1.69 seconds
      cpu time            1.31 seconds


177
178  proc print data = summary2 label;
179  run;

NOTE: There were 435 observations read from the data set WORK.SUMMARY2.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.32 seconds
      cpu time            0.29 seconds


180
181  ods pdf close;
NOTE: ODS PDF printed 59 pages to

C:\Users\dsingh\Dropbox\Tamu\Stat_604\Homework\HW5\DSingh_HW05_output.pdf
```