

About the Authors

Andrew T. Kuligowski has been a SAS user for over 25 years – WELL over. Currently the Manager of CRM Data Infrastructure at HSN in St. Petersburg, he has augmented his professional coding experiences in the retail, media, insurance, and petrochemical fields by speaking at various SAS conferences and user group meetings. Andrew was conference chair of SAS Global Forum 2012 in Orlando, and co-chair for SESUG'97 in Jacksonville and Tennessee SAS Users Day in Knoxville. In his spare time, Andrew volunteers at the Florida Aquarium in Tampa.

Swati Agarwal has been a SAS user for over 13 years. Currently working as Sr Data Consultant in Optum, Eden Prairie MN, she has augmented her professional coding experiences in insurance and healthcare fields by speaking at various SAS conferences and user groups meetings. Swati is co-founder of MinnSUG (Minnesota SAS User Group) and section chair in several MWSUG and local SAS user groups.



Secrets of Efficient SAS® Coding Techniques

Andrew T. Kuligowski
Swati Agarwal



of Efficient SAS[®] Coding ques

T. Kuligowski
garwal

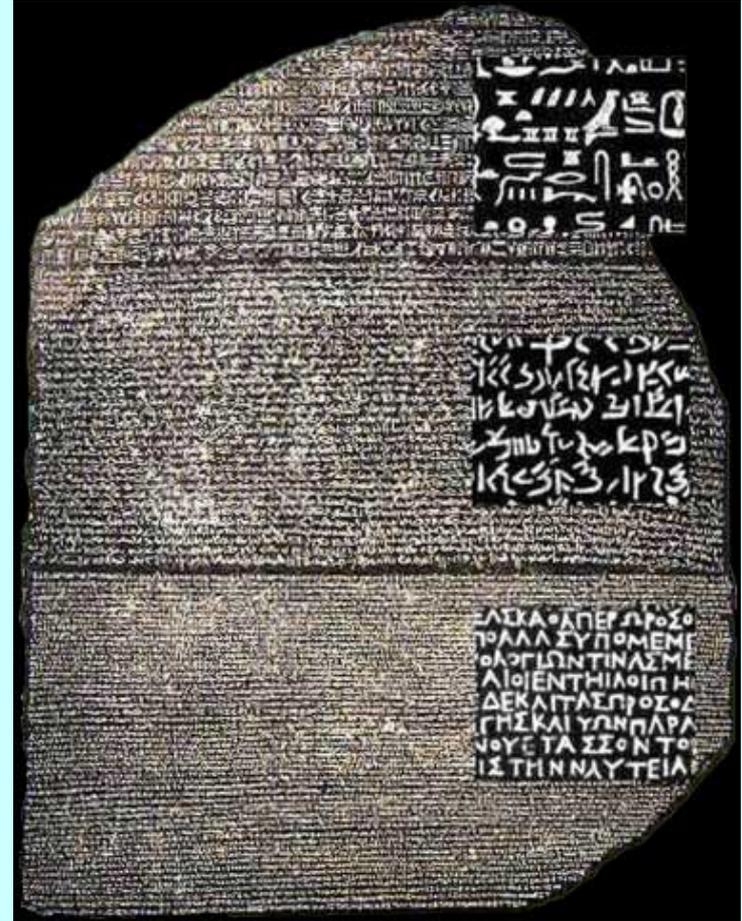
Learning French on a crash course ...

My favorite flowers:



Learning French on a crash course ...

Rosetta Stone



Learning French on a crash course ...

Audio CD:



À quelle distance se trouve la Tour Eiffel?
Savez-vous nager, monsieur?

Learning French on a crash course ...

Audio CD:



Où est la cathédrale Notre-Dame?

Learning French on a crash course ...

Merci de m'avoir invité dans votre belle ville.



What is Efficiency?

Dataset we are using for reference

The screenshot displays the SAS Enterprise Guide interface. The main window shows the 'The CONTENTS Procedure' for the dataset 'WORK.CSP_CLM_AFTER1'. The dataset properties are as follows:

Property	Value
Data Set Name	WORK.CSP_CLM_AFTER1
Member Type	DATA
Engine	V9
Created	04/12/2016 11:24:32
Last Modified	04/12/2016 11:24:32
Protection	
Data Set Type	
Label	
Data Representation	HP_UX_64, RS_6000_AIX_64, SOLARIS_64, HP_IA64
Encoding	latin1 - Western (ISO)
Observations	114382468
Variables	36
Indexes	0
Observation Length	304
Deleted Observations	0
Compressed	NO
Sorted	NO

Below the dataset properties, the 'Engine/Host Dependent Information' is shown:

Data Set Page Size	262144
Number of Data Set Pages	132849
First Data Page	1
Max Obs per Page	861
Obs in First Data Page	839
Number of Data Set Repairs	0
Filename	/arc_saswork/SAS_work7F7301B5012C_apspd0335/SAS_work9F9A01B5012C_apspd0335/csp_clm_after1.sas7bdat
Release Created	9.0401M3
Host Created	AIX
Inode Number	213965
Access Permission	rwx-rw-r--
Owner Name	saganw5
File Size	32GB
File Size (bytes)	34825830400

At the bottom of the main window, there is an 'Alphabetic List of Variables and Attributes' table with columns: #, Variable, Type, Len, Format, Informat, and Label.

Report last modified on 4/12/2016 11:33:02 AM.

The taskbar at the bottom shows the system tray with the date and time: 11:33 AM, Tuesday, 4/12/2016.

CODING TIP #1: When performing calculations make sure you execute only necessary variables and observations.

```
44   proc sort data = csp_clm_after1 out= csp_clm;  
45   by clcl_id;  
46   run;
```

NOTE: There were 114382468 observations read from the data set WORK.CSP_CLM_AFTER1.

NOTE: The data set WORK.CSP_CLM has **114,382,468** observations and 36 variables.

NOTE: PROCEDURE SORT used (Total process time):

```
real time      1:32.18  
user cpu time  40.39 seconds  
system cpu time 13.94 seconds
```

```
50   proc sort data = csp_clm_after1 out= csp_clm2;  
51   by clcl_id;  
52   run;
```

NOTE: There were 10000000 observations read from the data set WORK.CSP_CLM_AFTER1.

NOTE: The data set WORK.CSP_CLM2 has **10,000,000** observations and 36 variables.

NOTE: PROCEDURE SORT used (Total process time):

```
real time      7.10 seconds  
user cpu time  5.22 seconds  
system cpu time 1.34 seconds
```

```
56   proc sort data = csp_clm_after1 (keep = clcl_id PAID_DT PAID_AMT GRGR_ID SBSB_ID MEMBERNUM ) out= csp_clm3;  
57   by clcl_id;  
58   run;
```

NOTE: There were 114382468 observations read from the data set WORK.CSP_CLM_AFTER1.

NOTE: The data set WORK.CSP_CLM3 has **114,382,468** observations and 6 variables.

NOTE: PROCEDURE SORT used (Total process time):

```
real time      44.12 seconds  
user cpu time  32.71 seconds  
system cpu time 6.03 seconds
```

CODING TIP #1: When performing calculations make sure you execute only necessary variables and observations.

```
44 proc sort data = csp_clm_after1 out= csp_clm;  
45 by clcl_id;  
46 run;
```

NOTE: There were 114382468 observations read from the data set WORK.CSP_CLM_AFTER1.

NOTE: The data set WORK.CSP_CLM has **114,382,468** observations and 36 variables.

NOTE: PROCEDURE SORT used (Total process time):

```
real time      1:32.18  
user cpu time  40.39 seconds  
system cpu time 13.94 seconds
```

```
50 proc sort data = csp_clm_after1 out= csp_clm2;  
51 by clcl_id;  
52 run;
```

NOTE: There were 10000000 observations read from the data set WORK.CSP_CLM_AFTER1.

NOTE: The data set WORK.CSP_CLM2 has **10,000,000** observations and 36 variables.

NOTE: PROCEDURE SORT used (Total process time):

```
real time      7.10 seconds  
user cpu time   5.22 seconds  
system cpu time 1.34 seconds
```

```
56 proc sort data = csp_clm_after1 (keep = clcl_id PAID_DT PAID_AMT GRGR_ID SBSB_ID  
57 by clcl_id;  
58 run;
```

NOTE: There were 114382468 observations read from the data set WORK.CSP_CLM_AFTER1.

NOTE: The data set WORK.CSP_CLM3 has **114,382,468** observations and 6 variables.

NOTE: PROCEDURE SORT used (Total process time):

```
real time      44.12 seconds  
user cpu time  32.71 seconds  
system cpu time 6.03 seconds
```

Real time = wall clock time

System CPU time =
amount of time used by the
operating system on behalf
of SAS

User CPU time = amount
of time SAS used

CODING TIP #1: When performing calculations make sure you execute only necessary variables and observations.

```
44 proc sort data = csp_clm_after1 out= csp_clm;  
45 by clcl_id;  
46 run;
```

NOTE: There were **114,382,468 observations** read from the data set WORK.CSP_CLM_AFTER1.

NOTE: The data set WORK.CSP_CLM has 114382468 observations and 36 variables.

NOTE: PROCEDURE SORT used (Total process time):

```
real time      1:32.18  
user cpu time  40.39 seconds  
system cpu time 13.94 seconds
```

```
50 proc sort data = csp_clm_after1 out= csp_clm2;  
51 by clcl_id;  
52 run;
```

NOTE: There were **10,000,000 observations** read from the data set WORK.CSP_CLM_AFTER1.

NOTE: The data set WORK.CSP_CLM2 has 10000000 observations and 36 variables.

NOTE: PROCEDURE SORT used (Total process time):

```
real time      7.10 seconds  
user cpu time  5.22 seconds  
system cpu time 1.34 seconds
```

```
56 proc sort data = csp_clm_after1 (keep = clcl_id PAID_DT  
PAID_AMT GRGR_ID SBSB_ID MEMBERNUM ) out= csp_clm3;  
57 by clcl_id;  
58 run;
```

NOTE: There were **114,382,468 observations** read from the data set WORK.CSP_CLM_AFTER1.

NOTE: The data set WORK.CSP_CLM3 has 114382468 observations and 6 variables.

NOTE: PROCEDURE SORT used (Total process time):

```
real time      44.12 seconds  
user cpu time  32.71 seconds  
system cpu time 6.03 seconds
```

CODING TIP #5: Use the IN operator rather than logical OR operators.

```
64   data csp;  
65   set csp_clm_after1;  
66   where grgr_id in ('KSKCMD','TXSTPL','WAHLOP');  
67   run;
```

NOTE: There were 16971440 observations read from the data set WORK.CSP_CLM_AFTER1.

```
WHERE grgr_id in ('KSKCMD', 'TXSTPL', 'WAHLOP');
```

NOTE: The data set WORK.CSP has 16971440 observations and 36 variables.

```
real time      17.90 seconds  
user cpu time   4.59 seconds  
system cpu time 2.92 seconds
```

```
69   data csp1;  
70   set csp_clm_after1 (where = (grgr_id in ('KSKCMD','TXSTPL','WAHLOP')));  
71   run;
```

NOTE: There were 16971440 observations read from the data set WORK.CSP_CLM_AFTER1.

```
WHERE grgr_id in ('KSKCMD', 'TXSTPL', 'WAHLOP');
```

NOTE: The data set WORK.CSP1 has 16971440 observations and 36 variables.

NOTE: DATA statement used (Total process time):

```
real time      19.73 seconds  
user cpu time   4.60 seconds  
system cpu time 2.82 seconds
```

CODING TIP #5: Use the IN operator rather than logical OR operators.

```
73    data csp2;  
74    set csp_clm_after1;  
75    where grgr_id = 'KSKCMD' or grgr_id = 'TXSTPL' or grgr_id = 'WAHLOP';  
76    run;
```

NOTE: There were 16971440 observations read from the data set WORK.CSP_CLM_AFTER1.

```
WHERE grgr_id in ('KSKCMD', 'TXSTPL', 'WAHLOP');
```

NOTE: The data set WORK.CSP2 has 16971440 observations and 36 variables.

NOTE: DATA statement used (Total process time):

```
real time      19.80 seconds  
user cpu time   4.59 seconds  
system cpu time 2.88 seconds
```

```
78    proc sql;  
79    create table csp3 as  
80    (select *  
81    from csp_clm_after1 where grgr_id in ('KSKCMD','TXSTPL','WAHLOP'));
```

NOTE: Table WORK.CSP3 created, with 16971440 rows and 36 columns.

```
82    quit;
```

NOTE: PROCEDURE SQL used (Total process time):

```
real time      16.62 seconds  
user cpu time   5.00 seconds  
system cpu time 2.98 seconds
```

CODING TIP #5: Use the IN operator rather than logical OR operators.

<pre>64 data csp; 65 set csp_clm_after1; 66 where grgr_id in ('KSKCMD','TXSTPL','WAHLOP'); 67 run;</pre>	<p>NOTE: There were 16,971,440 observations read from the data set WORK.CSP_CLM_AFTER1.</p> <pre>WHERE grgr_id in ('KSKCMD', 'TXSTPL', 'WAHLOP');</pre> <p>NOTE: The data set WORK.CSP has 16971440 observations and 36 variables.</p> <pre>real time 17.90 seconds user cpu time 4.59 seconds system cpu time 2.92 seconds</pre>
<pre>69 data csp1; 70 set csp_clm_after1 (where = (grgr_id in ('KSKCMD','TXSTPL','WAHLOP'))); 71 run;</pre>	<p>NOTE: There were 16,971,440 observations read from the data set WORK.CSP_CLM_AFTER1.</p> <pre>WHERE grgr_id in ('KSKCMD', 'TXSTPL', 'WAHLOP');</pre> <p>NOTE: The data set WORK.CSP1 has 16971440 observations and 36 variables.</p> <p>NOTE: DATA statement used (Total process time):</p> <pre>real time 19.73 seconds user cpu time 4.60 seconds system cpu time 2.82 seconds</pre>
<pre>73 data csp2; 74 set csp_clm_after1; 75 where grgr_id = 'KSKCMD' or grgr_id = 'TXSTPL' or grgr_id = 'WAHLOP'; 76 run;</pre>	<p>NOTE: There were 16,971,440 observations read from the data set WORK.CSP_CLM_AFTER1.</p> <pre>WHERE grgr_id in ('KSKCMD', 'TXSTPL', 'WAHLOP');</pre> <p>NOTE: The data set WORK.CSP2 has 16971440 observations and 36 variables.</p> <p>NOTE: DATA statement used (Total process time):</p> <pre>real time 19.80 seconds user cpu time 4.59 seconds system cpu time 2.88 seconds</pre>
<pre>78 proc sql; 79 create table csp3 as 80 (select * 81 from csp_clm_after1 where grgr_id in ('KSKCMD','TXSTPL','WAHLOP'));</pre>	<p>NOTE: Table WORK.CSP3 created, with 16,971,440 rows and 36 columns.</p> <pre>82 quit;</pre> <p>NOTE: PROCEDURE SQL used (Total process time):</p> <pre>real time 16.62 seconds user cpu time 5.00 seconds system cpu time 2.98 seconds</pre>

CODING TIP #6: Use a series of If-THEN clauses rather than compound expressions with AND

```
88 data grgr;  
89 set csp_clm_after1;  
90 if business_line = 'MCD' and cobtype = 'NONCOB' and grgr_id = 'NJFAMCAR' then output;  
91 run;
```

NOTE: There were 114382468 observations read from the data set WORK.CSP_CLM_AFTER1.

NOTE: The data set WORK.GRGR has 13839212 observations and 36 variables.

NOTE: DATA statement used (Total process time):

```
real time      23.38 seconds  
user cpu time   9.28 seconds  
system cpu time 2.82 seconds
```

```
93 data grgr1;  
94 set csp_clm_after1;  
95 Where business_line = 'MCD' and cobtype = 'NONCOB' and grgr_id = 'NJFAMCAR' ;  
96 run;
```

NOTE: There were 13839212 observations read from the data set WORK.CSP_CLM_AFTER1.

WHERE (business_line='MCD') and (cobtype='NONCOB') and (grgr_id='NJFAMCAR');

NOTE: The data set WORK.GRGR1 has 13839212 observations and 36 variables.

NOTE: DATA statement used (Total process time):

```
real time      18.90 seconds  
user cpu time   6.23 seconds  
system cpu time 2.84 seconds
```

CODING TIP #6: Use a series of If-THEN clauses rather than compound expressions with AND

```
103  proc sql;
104  create table grgr2 as
105  (select *
106  from csp_clm_after1
107  where
108  business_line = 'MCD'
109  and
110  cobtype = 'NONCOB'
111  and
112  grgr_id = 'NJFAMCAR');
```

NOTE: Table WORK.GRGR2 created, with 13839212 rows and 36 columns.

NOTE: PROCEDURE SQL used (Total process time):

```
real time      19.33 seconds
user cpu time   6.53 seconds
system cpu time 2.86 seconds
```

CODING TIP #6: Use a series of If-THEN clauses rather than compound expressions with AND

```
88 data grgr;  
89 set csp_clm_after1;  
90 if business_line = 'MCD' and cobtype = 'NONCOB' and grgr_id =  
'NJFAMCAR' then output;  
91 run;
```

NOTE: There were 114382468 observations read from the data set WORK.CSP_CLM_AFTER1.
NOTE: The data set WORK.GRGR has 13839212 observations and 36 variables.
NOTE: DATA statement used (Total process time):
real time 23.38 seconds
user cpu time 9.28 seconds
system cpu time 2.82 seconds

```
93 data grgr1;  
94 set csp_clm_after1;  
95 where  
96 business_line = 'MCD'  
97 and  
98 cobtype = 'NONCOB'  
99 and  
100 grgr_id = 'NJFAMCAR' ;  
101 run;
```

NOTE: There were 13839212 observations read from the data set WORK.CSP_CLM_AFTER1.
WHERE (business_line='MCD') and (cobtype='NONCOB') and (grgr_id='NJFAMCAR');
NOTE: The data set WORK.GRGR1 has 13839212 observations and 36 variables.
NOTE: DATA statement used (Total process time):
real time 18.90 seconds
user cpu time 6.23 seconds
system cpu time 2.84 seconds

```
103 proc sql;  
104 create table grgr2 as  
105 (select *  
106 from csp_clm_after1  
107 where  
108 business_line = 'MCD'  
109 and  
110 cobtype = 'NONCOB'  
111 and  
112 grgr_id = 'NJFAMCAR');
```

NOTE: Table WORK.GRGR2 created, with 13839212 rows and 36 columns.
NOTE: PROCEDURE SQL used (Total process time):
real time 19.33 seconds
user cpu time 6.53 seconds
system cpu time 2.86 seconds

CODING TIP #8: Read only the fields you need.

```
119  data test;  
120  set csp_clm_after1;  
121  keep clcl_id PAID_DT PAID_AMT GRGR_ID SBSB_ID MEMBERNUM;  
122  run;
```

NOTE: There were 114382468 observations read from the data set WORK.CSP_CLM_AFTER1.

NOTE: The data set WORK.TEST has 114382468 observations and 6 variables.

NOTE: DATA statement used (Total process time):

```
real time      26.49 seconds  
user cpu time   9.24 seconds  
system cpu time 3.23 seconds
```

```
124  data test1;  
125  set csp_clm_after1 (keep = clcl_id PAID_DT PAID_AMT GRGR_ID SBSB_ID MEMBERNUM );  
126  run;
```

NOTE: There were 114382468 observations read from the data set WORK.CSP_CLM_AFTER1.

NOTE: The data set WORK.TEST1 has 114382468 observations and 6 variables.

NOTE: DATA statement used (Total process time):

```
real time      25.72 seconds  
user cpu time   9.20 seconds  
system cpu time 3.36 seconds
```

CODING TIP #8: Read only the fields you need.

```
128 data test2 (keep = clcl_id PAID_DT PAID_AMT GRGR_ID SBSB_ID MEMBERNUM );  
129 set csp_clm_after1;  
130 run;
```

NOTE: There were 114382468 observations read from the data set WORK.CSP_CLM_AFTER1.

NOTE: The data set WORK.TEST2 has 114382468 observations and 6 variables.

NOTE: DATA statement used (Total process time):

```
real time      26.32 seconds  
user cpu time   9.24 seconds  
system cpu time 3.21 seconds
```

CODING TIP #8: Read only the fields you need.

```
119 data test;  
120 set csp_clm_after1;  
121 keep clcl_id PAID_DT PAID_AMT GRGR_ID SBSB_ID MEMBERNUM;  
122 run;
```

NOTE: There were 114382468 observations read from the data set WORK.CSP_CLM_AFTER1.
NOTE: The data set WORK.TEST has 114382468 observations and 6 variables.
NOTE: DATA statement used (Total process time):

real time	26.49 seconds
user cpu time	9.24 seconds
system cpu time	3.23 seconds

```
124 data test1;  
125 set csp_clm_after1 (keep = clcl_id PAID_DT PAID_AMT GRGR_ID  
SBSB_ID MEMBERNUM );  
126 run;
```

NOTE: There were 114382468 observations read from the data set WORK.CSP_CLM_AFTER1.
NOTE: The data set WORK.TEST1 has 114382468 observations and 6 variables.
NOTE: DATA statement used (Total process time):

real time	25.72 seconds
user cpu time	9.20 seconds
system cpu time	3.36 seconds

```
128 data test2 (keep = clcl_id PAID_DT PAID_AMT GRGR_ID SBSB_ID  
MEMBERNUM );  
129 set csp_clm_after1;  
130 run;
```

NOTE: There were 114382468 observations read from the data set WORK.CSP_CLM_AFTER1.
NOTE: The data set WORK.TEST2 has 114382468 observations and 6 variables.
NOTE: DATA statement used (Total process time):

real time	26.32 seconds
user cpu time	9.24 seconds
system cpu time	3.21 seconds

CODING TIP #12: Use the LENGTH statement to reduce storage space for variables in SAS data sets

```
31   data test3;  
32   set test2;  
33   temp_var = GRGR_ID || MEMBERNUM || ccl_id ; run;
```

NOTE: There were 114382468 observations read from the data set WORK.TEST2.

NOTE: The data set WORK.TEST3 has 114382468 observations and 7 variables.

NOTE: DATA statement used (Total process time):

```
real time      29.41 seconds  
user cpu time  11.96 seconds  
system cpu time 2.39 seconds
```

```
38   data test3a;  
39   set test2;  
40   length temp_var $40;  
41   temp_var = GRGR_ID || MEMBERNUM || ccl_id ;  
42   run;
```

NOTE: There were 114382468 observations read from the data set WORK.TEST2.

NOTE: The data set WORK.TEST3A has 114382468 observations and 7 variables.

NOTE: DATA statement used (Total process time):

```
real time      26.22 seconds  
user cpu time  11.67 seconds  
system cpu time 2.03 seconds
```

CODING TIP #12: Use the LENGTH statement to reduce storage space for variables in SAS data sets

```
31 data test3;  
32 set test2;  
33 temp_var = GRGR_ID || MEMBERNUM || clcl_id ; run;
```

NOTE: There were 114382468 observations read from the data set WORK.TEST2.
NOTE: The data set WORK.TEST3 has 114382468 observations and 7 variables.
NOTE: DATA statement used (Total process time):
real time 29.41 seconds
user cpu time 11.96 seconds
system cpu time 2.39 seconds

```
38 data test3a;  
39 set test2;  
40 length temp_var $40;  
41 temp_var = GRGR_ID || MEMBERNUM || clcl_id ;  
42. run;
```

NOTE: There were 114382468 observations read from the data set WORK.TEST2.
NOTE: The data set WORK.TEST3A has 114382468 observations and 7 variables.
NOTE: DATA statement used (Total process time):
real time 26.22 seconds
user cpu time 11.67 seconds
system cpu time 2.03 seconds

