ABSTRACT:

Have you ever been asked to compare new datasets to old datasets while transfers of data occur several times during the project life time, yet the dataset names, to compare are different and each dataset has different key variables? SAS® macro facility along with metadata files with rules on how to compare can be used for orchestration of this process. The Rules metadata file contains information about filenames (old & new), key variables to be used to compare each dataset and variables not to compare for each dataset.

Focus on the metadata file will result in the tool providing meaningful output to review. This paper will discuss the process and identify how it provides confidence in the output.

INTRODUCTION:

Data comparison is an important aspect in any project. This paper will discuss a macro to produce a single report after comparing datasets regardless of datasets being compared.

MACRO OVERVIEW:

The COMPARE macro has the following steps.

1.) The macro imports the Excel sheet.
   a.) The Excel consists of following details as shown in fig 1.
      I. The by variables to be used to compare the datasets.
      II. Drop variables that we don’t want to compare in each dataset.
      III. If the dataset is to be compared or not.

2.) It creates macro variables of the dataset names, by variables, drop variables using call symput function.

3.) The do loop is used to process all the datasets at a time.

Fig1: Excel sheet with specifications.
DATA:
The libname statements are used to define the libraries.
**First Data Tranfer;
libname trans1 "C:\nesug\datatransfer1";
**Second Data Transfer;
libname trans2 "C:\nesug\datatransfer2";

**Salary data of each empid;
data trans1.salary;
   input empid position :$15. salary;
   format empid z3. salary dollar10.;
datalines;
001 Sr.Manager    100000
002 Manager       90000
003 Teamlead      80000
004 Sr.programmer 75000
005 programmer    60000
;
run;
data trans2.salary1;
   input empid position :$15. salary;
   format empid z3. salary dollar10.;
datalines;
001 Sr.Manager    100000
002 Manager       90000
003 Teamlead      80000
004 Sr.programmer 75000
005 programmer    65000
;
run;
**Bonus data of each empid;
data trans1.bonus;
   input empid position :$15. salary percent;
   format empid z3. salary dollar10.;
   bonus=(percent/100)*salary;
datalines;
001 Sr.Manager    100000 20
002 Manager       90000  15
003 Teamlead      80000  10
004 Sr.programmer 75000   4
005 programmer    60000   3
;
run;
data trans2.bonus1;
   input empid position :$15. salary percent;
   format empid z3. salary dollar10.;
   bonus=(percent/100)*salary;
datalines;
001 Sr.Manager    100000 20
002 Manager       90000  15
003 Teamlead      80000  10
004 Sr.programmer 75000   4
005 programmer    65000   3
;
run;
**Shares data of each empid;
data trans1.shares;
   input empid position :$15. salary percent shares;
   format empid z3. salary dollar10.;
   bonus=(percent/100)*salary;
datalines;
001 Sr.Manager    100000 20 10000
002 Manager       90000 15 5000
003 Teamlead      80000 10
004 Sr.programmer 75000   4
005 programmer    65000   3
;
run;
<table>
<thead>
<tr>
<th>Emp ID</th>
<th>Position</th>
<th>Salary</th>
<th>Percent</th>
<th>Shares</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>Sr. Manager</td>
<td>100000</td>
<td>20</td>
<td>10000</td>
</tr>
<tr>
<td>002</td>
<td>Manager</td>
<td>90000</td>
<td>15</td>
<td>5000</td>
</tr>
<tr>
<td>003</td>
<td>Teamlead</td>
<td>80000</td>
<td>10</td>
<td>3000</td>
</tr>
<tr>
<td>004</td>
<td>Sr. Programmer</td>
<td>75000</td>
<td>4</td>
<td>1000</td>
</tr>
<tr>
<td>005</td>
<td>Programmer</td>
<td>65000</td>
<td>3</td>
<td>500</td>
</tr>
</tbody>
</table>

MACRO:

```bash
%macro compare(lib1, lib2, xls);
libname lib1 "&&lib1.";
libname lib2 "&&lib2.";
**Import Specifications for each dataset;
proc import datafile = "&&xls"
  out = qcl
    dbms = excel replace;
    getnames = yes;
    sheet = compare;
run;

**Create macro variables;
data qc;
set qcl(where=(dataset1^=' ' and qc='Y'));
retain i;
  i + 1;
  call symput("dat" || trim(left(put(i,3.))), trim(dataset1));
  call symput("dat2" || trim(left(put(i,3.))), trim(dataset2));
  call symput("byvar" || trim(left(put(i,3.))), trim(Byvar));
  call symput("qc" || trim(left(put(i,3.))), trim(QC));
  call symput("dropvar" || trim(left(put(i,3.))), trim(dropvar));
run;

***Print the total number of datasets being compared to log;
%put Total number of datasets=&tot;

**Output the results to Rtf file;
options orientation=landscape;
ods listing close;
  ods noresults;
  ods rtf file="C:\nesug\compare.rtf";

**Run all the datasets at once;
%do i=1 %to &tot;
  %if %upcase(&dropvar&i.)=NONE %then %do;
    **Sort using the by variables before comparing and output to another dataset ;
```
%end; %else %do;
**Sort using the by variables before comparing and output to another dataset ;
proc sort data=lib1.&dat1 out=base&i.(drop=&dropvar&i.);
   by &byvar&i.;
run;
proc sort data=lib2.&dat2'i. out=compare&i.(drop=&dropvar&i.);
   by &byvar&i.;
run;
%end; ***Comparing the datasets;
title1 "Comparison is done for data in &lib1. with &lib2. ";
title2 "Base: &dat1 Compare: &dat2' ";
proc compare data=base&i. compare=compare&i. ;
   id &byvar&i.;
run;
%end;
ods rtf close;
ods results;
ods listing;
%MEND;
**Run Macro;
%compare(%str(C:\nesug\datatransfer1),%str(C:\nesug\datatransfer2),
%str(C:\nesug\compare.xls));

OUTPUT:

\[\text{Comparison is done for data in datatransfer1 with datatransfer2} \]
\[\text{Base: salary Compare: salary} \]

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
\textbf{Variable} & \textbf{Base} & \textbf{Compare} \\
\hline
\text{ID} & 1 & 5 \\
\text{Salary} & 1000 & 1500 \\
\hline
\end{tabular}
\end{table}

4
CONCLUSIONS:

As mentioned in above method, this compare macro is designed to make your life easier. You can use this code to use it over and over again.

We can also use this to change to merge statement than using proc compare.

REFERENCES:


ACKNOWLEDGMENTS:

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

Many thanks to Daphne Ewing at Auxilium Pharmaceuticals for her support and encouragement to complete this paper.

CONTACT INFORMATION:

Your comments and questions are valued and encouraged. Contact the author at:

   Sandeep Kottam
   Remx IT
   201 Heritage In
   Exton, PA-19341
   Phone: 605-691-3274
   kottamsandeep@gmail.com