Pre Data Quality Macro

PREDQ Macro:

Description	This macro is used to produce a summary report containing number of non-missing records in each variable for two datasets and highlight the following: 1) Decline in number of non-missing values in a variable. 2) Any format changes for a variable in two datasets. 3) New or dropped variables. This produces output in excel format or in "Results – SAS Report" tab (in SAS EG). >>Orange implies 0% to 10% drop in number of records >>Red implies 10% to 100% drop in number of records >>Yellow on "Old vs New Data Type" column cell implies data type change >>Light Green on "Old vs New Data Type" column cell implies data type is similar but format is different "Old vs New Data Type" will not appear if NEW_DTSET is blank. **Script Location : G:\dqmacro\predq.sas**

Syntax	%PREDQ(DOMAIN=
	    ,DB=
	    ,NEW_DTSET=
	    ,OUTPUT_DIR=);

Parameters	%PREDQ(DOMAIN = Name of the library like health, social etc..
	    ,DB = Database name to compare to. It will automatically pull last member in the cluster unless providing memnum in both DB & NEW_DTSET parameters.
	    ,NEW_DTSET=New Database/cluster member. Keep this field blank if both the databases to compare are in same cluster. Provide memnum/dataset name to compare specific member.
	    ,OUTPUT_DIR=Location of output.

Example	1) When both comparing datasets are in same cluster: %PREDQ(DOMAIN=HEALTH, DB=MHMHLTH_MHID_2010APR, NEW_DTSET=, OUTPUT_DIR=G:\dmusers\username); 2) When both comparing datasets are in different cluster: %PREDQ(DOMAIN=common, DB=conc_2006, NEW_DTSET=conc_2011, OUTPUT_DIR=G:\dmusers\username); 3) When want to compare specific members in same/different cluster: %PREDQ(DOMAIN=REGISTRY, DB=VSA_MMDF_2000JAN(MEMNUM=2), NEW_DTSET=VSA_MMDF_2000JAN(MEMNUM=3), OUTPUT_DIR=G:\dmusers\username);
Sample Output	Orange implies 0% to 10% drop in number of records Red implies 10% to 100% drop in number of records Yellow on "Old vs New Data Type" column implies data type change Light Green on "Old vs New Data Type" column implies data type is similar but format is different *Latest Cluster Member: conc_2011(MEMNUM=1)* *Previous Cluster member: conc_2006(MEMNUM=1)* <table><tr><th>Variables In Cluster</th><th>Previous Cluster Member Count</th><th>Latest Cluster Member Count</th><th>Percent Change</th><th>Previous vs Latest Data Type</th></tr><tr><td>DAUID</td><td>.</td><td>2179</td><td>Variable not in Previous Cluster</td><td>Char(8)</td></tr><tr><td>CC_F_POP</td><td>54529</td><td>2176</td><td>-96.01%</td><td>Num(8)(BEST12) VS Num(8)</td></tr><tr><td>DB_IR2011</td><td>.</td><td>2179</td><td>Variable not in Previous Cluster</td><td>Char(1)</td></tr><tr><td>CTNAME</td><td>54626</td><td>2179</td><td>-96.01%</td><td>Char(9)($9) VS Char(7)</td></tr><tr><td>CTCODE</td><td>.</td><td>2179</td><td>Variable not in Previous Cluster</td><td>Char(4)</td></tr><tr><td>DA_UID</td><td>54626</td><td>.</td><td>Variable not in Latest Cluster</td><td>Char(10)($10)</td></tr></table>