PhilaSUG Winter 2017 Meeting

A Careful Approach when merging datasets - while sub setting using “WHERE or IF Statements”

PRESENTED BY

KAMALA B. MADAVARAPU, MS

Table of Contents

Abstract

- This short talk is intended to present how different the results could be and how data processes when doing a data merge while using WHERE or IF statements. And a trick to visualize the data override when merging.

My Background

- Education:
  - ✓ Masters: Analytical Chemistry, Governors State University, Illinois.
  - ✓ Post Graduate Diploma in Computer Applications, Creative Soft, India.
  - ✓ Bachelors: Pharmacy, Dr. M.G.R Medical University, India.
- I'm a certified Base SAS Programmer, working with SAS Technologies since 2008, Fundamentally Using it for data processing (ETL), analysis and Reporting.
- Into Consulting most of my career.
- Worked for Educational, Clinical Research, Health care and Financial domain clients.
- Currently working in Enterprise Data Management area for a Financial Client based in Wilmington, DE.
- This is my first SUG presentation.

Let's say we have two Individual Datasets

```
data one;
input  wave Channel$ percent_population ;
datalines;
1 DM 93
2 DM 92
3 DM 71
4 DM 99
;
run;
```

	wave	Channel	percent_population
1	1	DM	93
2	2	DM	92
3	3	DM	71
4	4	DM	99

```
data two;
input  wave Channel$ percent_population ;
datalines;
1 DM 93
2 DM 82
3 DM 71
4 DM 89
;
run;
```

	wave	Channel	percent_population
1	1	DM	93
2	2	DM	82
3	3	DM	71
4	4	DM	89

Want to merge them ? You have two Options

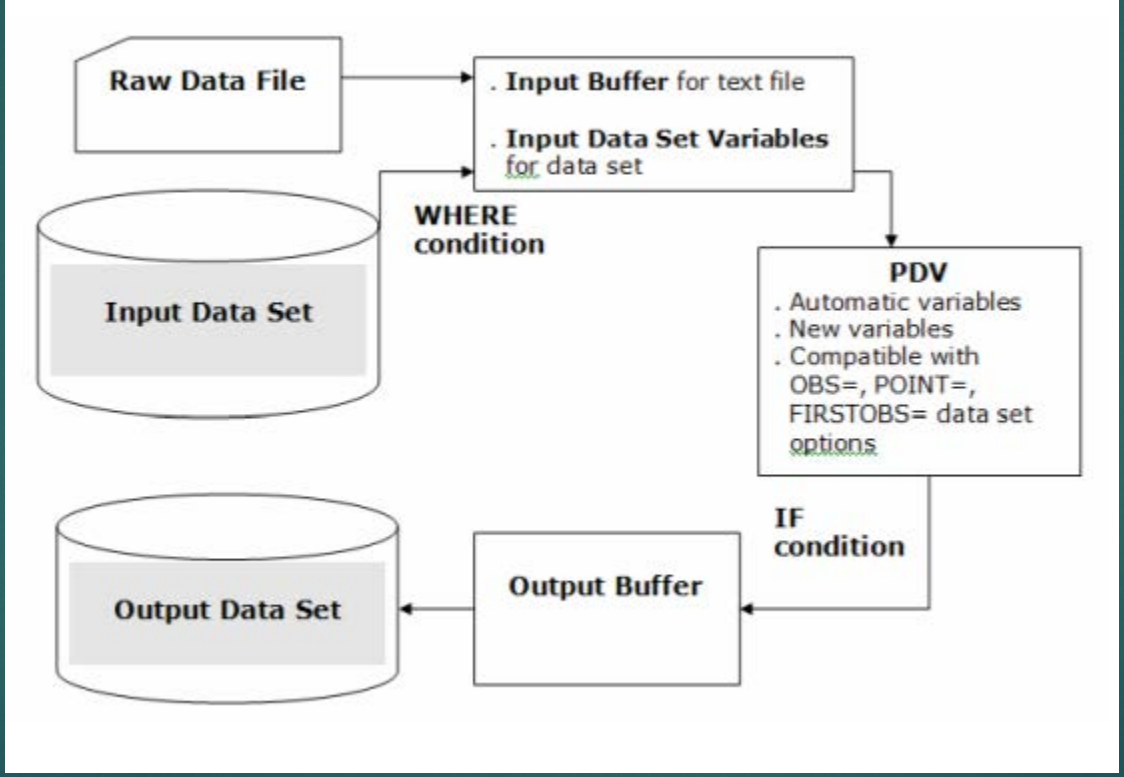
```
data use_where;
merge one two;
by wave;
where percent_population > 90;
run;
```

	wave	Channel	percent_population
1	1	DM	93
2	2	DM	92
3	4	DM	99

```
data use_if;
merge one two;
by wave;
if percent_population > 90;
run;
```

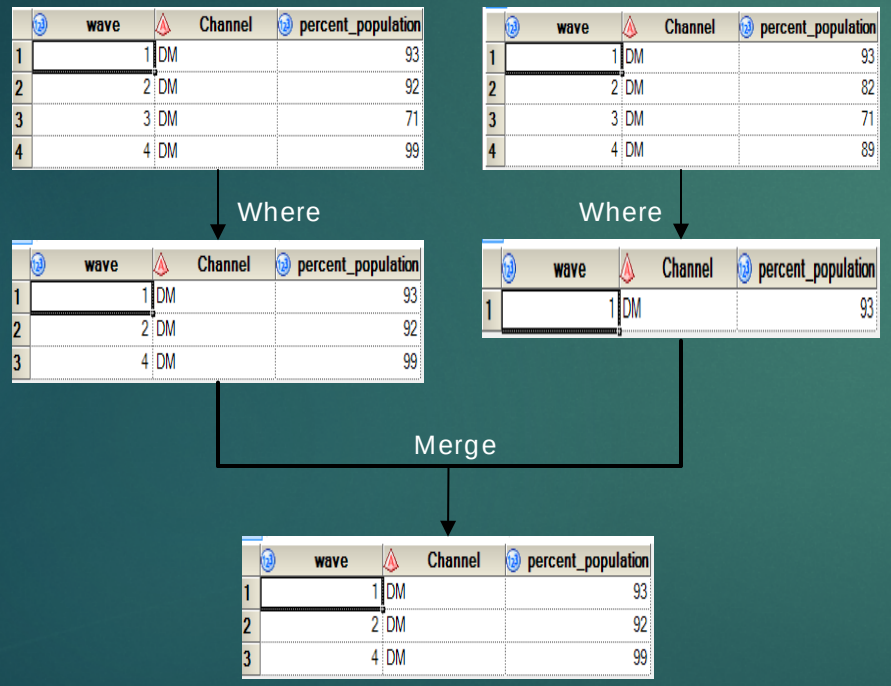
	wave	Channel	percent_population
1	1	DM	93

Why the difference? What is the Reason ?

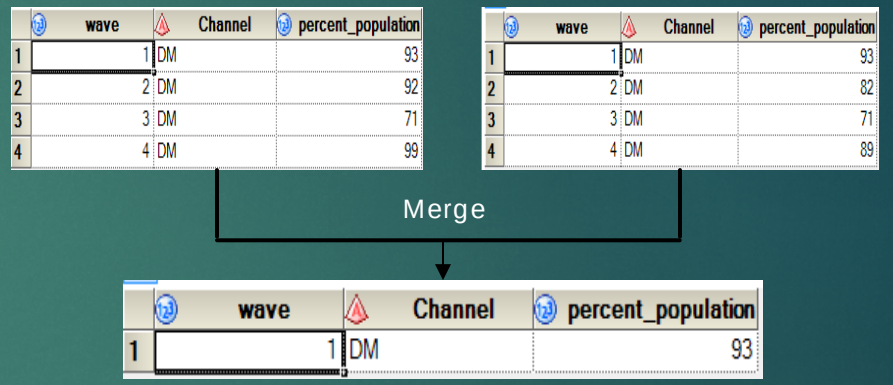


Let us see what actually happened in our datasets

- With <Where> statement



- With <If> statement



Points to Remember

- “Where” statement subsets data before merging.
- “If” statement subsets data after merging.

- In common, we want to apply IF statement after merging the datasets

How to Visualize it ?

- Use MSGLEVEL=System Option
- Syntax

  MSGLEVEL = I;
- Description

SAS writes additional informative messages to SAS Log pertaining to error messages, warnings, merge processing (that includes data overrides), index usage etc..

Visualize on SAS Log

```
55          data use_where;
56          merge one two;
57          by wave;
58          where poppercent > 90;
59          run;

INFO: The variable Channel on data set WORK.ONE will be overwritten by data set WORK.TWO.
INFO: The variable poppercent on data set WORK.ONE will be overwritten by data set WORK.TWO.
NOTE: Compression was disabled for data set WORK.USE_WHERE because compression overhead would increase
NOTE: There were 3 observations read from the data set WORK.ONE.
      WHERE poppercent>90;
NOTE: There were 1 observations read from the data set WORK.TWO.
      WHERE poppercent>90;
NOTE: The data set WORK.USE_WHERE has 3 observations and 3 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds
```

```
62          data use_if;
63          merge one two;
64          by wave;
65          if poppercent > 90;
66          run;

INFO: The variable Channel on data set WORK.ONE will be overwritten by data set WORK.TWO.
INFO: The variable poppercent on data set WORK.ONE will be overwritten by data set WORK.TWO.
NOTE: Compression was disabled for data set WORK.USE_IF because compression overhead would increase
NOTE: There were 4 observations read from the data set WORK.ONE.
NOTE: There were 4 observations read from the data set WORK.TWO.
NOTE: The data set WORK.USE_IF has 1 observations and 3 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
```

References

- http://www2.sas.com/proceedings/forum2007/213-2007.pdf
- http://support.sas.com/kb/24/286.html
- http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a000279149.htm

Email: kamala.madavarapu@gmail.com
Phone: 571-205-9084


thank
you