

SAS® GLOBAL FORUM 2017

April 2 - 5 | Orlando, FL

Quick Results with PROC SQL

A Hands-on Workshop by

Kirk Paul Lafler

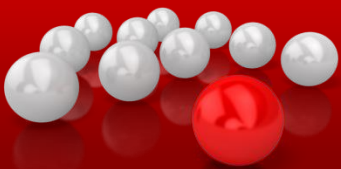
USERS PROGRAM



Copyright © 1992-2017 by
Kirk Paul Lafler and Software Intelligence Corporation.
All rights reserved.

SAS is the registered trademark of SAS Institute Inc., Cary, NC,
USA.

All other company and product names mentioned are used for
identification purposes only and may be trademarks of their
respective owners.



Workshop Objectives

Illustrate useful code examples to help SAS users:

- **Construct Simple Queries to Retrieve, Subset, and Search**
- **Summarize Data with Functions**
- **Explore Data Relationships with Joins**

Tables Used in Exercises

	Title	Length	Category	Year	Studio	Rating
1	Brave Heart	177	Action Adventure	1995	Paramount Pictures	R
2	Casablanca	103	Drama	1942	MGM / UA	PG
3	Christmas Vacation	97	Comedy	1989	Warner Brothers	PG-13
4	Coming to America	116	Comedy	1988	Paramount Pictures	R
5	Dracula	130	Horror	1993	Columbia TriStar	R
6	Dressed to Kill	105	Drama Mysteries	1980	Filmways Pictures	R
7	Forrest Gump	142	Drama	1994	Paramount Pictures	PG-13
8	Ghost	127	Drama Romance	1990	Paramount Pictures	PG-13
9	Jaws	125	Action Adventure	1975	Universal Studios	PG
10	Jurassic Park	127	Action	1993	Universal Pictures	PG-13
11	Lethal Weapon	110	Action Cops & Robber	1987	Warner Brothers	R
12	Michael	106	Drama	1997	Warner Brothers	PG-13
13	National Lampoon's Vacation	98	Comedy	1983	Warner Brothers	PG-13
14	Poltergeist	115	Horror	1982	MGM / UA	PG
15	Rocky	120	Action Adventure	1976	MGM / UA	PG
16	Scarface	170	Action Cops & Robber	1983	Universal Studios	R
17	Silence of the Lambs	118	Drama Suspense	1991	Orion	R
18	Star Wars	124	Action Sci-Fi	1977	Lucas Film Ltd	PG
19	The Hunt for Red October	135	Action Adventure	1989	Paramount Pictures	PG
20	The Terminator	108	Action Sci-Fi	1984	Live Entertainment	R
21	The Wizard of Oz	101	Adventure	1939	MGM / UA	G
22	Titanic	194	Drama Romance	1997	Paramount Pictures	PG-13

Movies

Actors

	Title	Actor_Leading	Actor_Supporting
1	Brave Heart	Mel Gibson	Sophie Marceau
2	Christmas Vacation	Chevy Chase	Beverly D'Angelo
3	Coming to America	Eddie Murphy	Arsenio Hall
4	Forrest Gump	Tom Hanks	Sally Field
5	Ghost	Patrick Swayze	Demi Moore
6	Lethal Weapon	Mel Gibson	Danny Glover
7	Michael	John Travolta	Andie MacDowell
8	National Lampoon's Vacation	Chevy Chase	Beverly D'Angelo
9	Rocky	Sylvester Stallone	Talia Shire
10	Silence of the Lambs	Anthony Hopkins	Jodie Foster
11	The Hunt for Red October	Sean Connery	Alec Baldwin
12	The Terminator	Arnold Schwarzenegger	Michael Biehn
13	Titanic	Leonardo DiCaprio	Kate Winslet


Introduction to PROC SQL

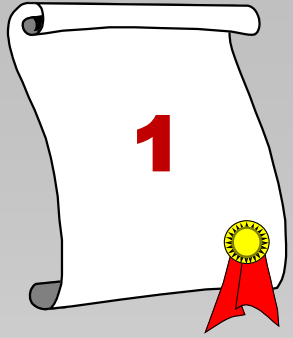


SQL – A Universal Database Language

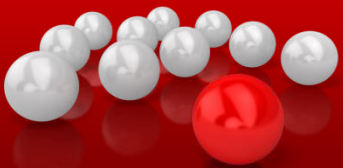
- PROC SQL works with tables of data
- Portable to other platforms and vendor's SQL implementation
- Universal (ANSI) language consisting of:
 - ✓ Statements / Clauses
 - ✓ Expressions
 - ✓ Operators
 - ✓ Summary functions

SQL and the SAS System



- **Part of the base product – no additional licenses**
 - **Can be used Interactively or Batch**
 - **Global statements such as TITLE, FOOTNOTE, and ODS may be used with PROC SQL**
 - **Databases are equivalent to libraries**
 - **Tables are equivalent to data sets**
 - **Rows are equivalent to observations**
 - **Columns are equivalent to variables**
- 
- A solid red bar with a diagonal cut across the bottom of the slide, starting from the left edge and extending towards the right.



Constructing PROC SQL Queries



Exercise #1

1. Click the Open folder  icon located at the top of the SAS Display Manager.
2. Select the SAS program named “Exer01” from the list.
3. The code should display in the SAS Editor.
4. Run the PROC SQL code by clicking the Submit  icon.
5. Let’s discuss the exercise and corresponding output.



Exercise #1 (continued)

An asterisk (*) or wildcard character in the SELECT statement displays all columns of data in the order they appear in the descriptor portion of the table.

```
PROC SQL ;  
  SELECT *  
    FROM SASUSER.MOVIES ;  
QUIT ;
```





Exercise #1 (continued)

The SAS System

Title	Length	Category	Year	Studio	Rating
Brave Heart	177	Action Adventure	1995	Paramount Pictures	R
Casablanca	103	Drama	1942	MGM / UA	PG
Christmas Vacation	97	Comedy	1989	Warner Brothers	PG-13
Coming to America	116	Comedy	1988	Paramount Pictures	R
Dracula	130	Horror	1993	Columbia TriStar	R
Dressed to Kill	105	Drama Mysteries	1980	Filmways Pictures	R
Forrest Gump	142	Drama	1994	Paramount Pictures	PG-13
Ghost	127	Drama Romance	1990	Paramount Pictures	PG-13
Jaws	125	Action Adventure	1975	Universal Studios	PG
Jurassic Park	127	Action	1993	Universal Pictures	PG-13
Lethal Weapon	110	Action Cops & Robber	1987	Warner Brothers	R
Michael	106	Drama	1997	Warner Brothers	PG-13
National Lampoon's Vacation	98	Comedy	1983	Warner Brothers	PG-13
Poltergeist	115	Horror	1982	MGM / UA	PG
Rocky	120	Action Adventure	1976	MGM / UA	PG
Scarface	170	Action Cops & Robber	1983	Universal Studios	R
Silence of the Lambs	118	Drama Suspense	1991	Orion	R
Star Wars	124	Action Sci-Fi	1977	Lucas Film Ltd	PG
The Hunt for Red October	135	Action Adventure	1989	Paramount Pictures	PG
The Terminator	108	Action Sci-Fi	1984	Live Entertainment	R
The Wizard of Oz	101	Adventure	1939	MGM / UA	G
Titanic	194	Drama Romance	1997	Paramount Pictures	PG-13

Exercise #2

1. Click the Open folder  icon located at the top of the SAS Display Manager.
2. Select the SAS program named “Exer02” from the list.
3. The code should display in the SAS Editor.
4. Run the PROC SQL code by clicking the Submit  icon.
5. Let’s discuss the exercise and corresponding output.



Exercise #2 (continued)

To display a select number of columns of data from a table, specify the name of each column separating each column name with a comma.

```
PROC SQL ;  
    SELECT title, rating  
        FROM SASUSER.MOVIES ;  
QUIT ;
```





Exercise #2 (continued)

The SAS System

<u>Title</u>	<u>Rating</u>
Brave Heart	R
Casablanca	PG
Christmas Vacation	PG-13
Coming to America	R
Dracula	R
Dressed to Kill	R
Forrest Gump	PG-13
Ghost	PG-13
Jaws	PG
Jurassic Park	PG-13
Lethal Weapon	R
Michael	PG-13
National Lampoon's Vacation	PG-13
Poltergeist	PG
Rocky	PG
Scarface	R
Silence of the Lambs	R
Star Wars	PG
The Hunt for Red October	PG
The Terminator	R
The Wizard of Oz	G
Titanic	PG-13

Exercise #3

1. Click the Open folder  icon located at the top of the SAS Display Manager.
2. Select the SAS program named “Exer03” from the list.
3. The code should display in the SAS Editor.
4. Run the PROC SQL code by clicking the Submit  icon.
5. Let’s discuss the exercise and corresponding output.



Exercise #3 (continued)

To order rows of data from a table in ascending order, specify the **ORDER BY** clause and the desired column name.

```
PROC SQL ;  
  SELECT title, rating  
    FROM SASUSER.MOVIES  
   ORDER BY rating;  
QUIT ;
```





Exercise #3 (continued)

The SAS System

<u>Title</u>	<u>Rating</u>
The Wizard of Oz	G
The Hunt for Red October	PG
Star Wars	PG
Poltergeist	PG
Jaws	PG
Rocky	PG
Casablanca	PG
Forrest Gump	PG-13
Christmas Vacation	PG-13
Michael	PG-13
National Lampoon's Vacation	PG-13
Jurassic Park	PG-13
Titanic	PG-13
Ghost	PG-13
Dressed to Kill	R
Lethal Weapon	R
Dracula	R
The Terminator	R
Brave Heart	R
Coming to America	R
Silence of the Lambs	R
Scarface	R

Exercise #4

1. Click the Open folder  icon located at the top of the SAS Display Manager.
2. Select the SAS program named “Exer04” from the list.
3. The code should display in the SAS Editor.
4. Run the PROC SQL code by clicking the Submit  icon.
5. Let’s discuss the exercise and corresponding output.



Exercise #4 (continued)

To remove duplicate rows from a table, the **DISTINCT** keyword can be specified prior to the column name that is being selected.

```
PROC SQL ;  
  SELECT DISTINCT rating  
    FROM SASUSER.MOVIES;  
QUIT ;
```



Exercise #4 (continued)

The SAS System

Rating



G

PG

PG-13

R

Exercise #5

1. Click the Open folder  icon located at the top of the SAS Display Manager.
2. Select the SAS program named “Exer05” from the list.
3. The code should display in the SAS Editor.
4. Run the PROC SQL code by clicking the Submit  icon.
5. Let’s discuss the exercise and corresponding output.



Exercise #5 (continued)

To select only those rows that match a desired value in a table, a LIKE operator is used with the percent '%' wildcard character in a WHERE clause.

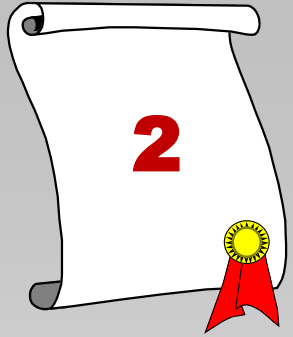
```
PROC SQL ;  
  SELECT title, category  
    FROM SASUSER.MOVIES  
   WHERE UPCASE(category) LIKE '%ACTION%';  
QUIT ;
```



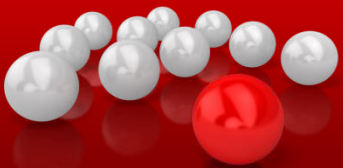
Exercise #5 (continued)

The SAS System

<u>Title</u>	<u>Category</u>
Brave Heart	Action Adventure
Jaws	Action Adventure
Jurassic Park	Action
Lethal Weapon	Action Cops & Robber
Rocky	Action Adventure
Scarface	Action Cops & Robber
Star Wars	Action Sci-Fi
The Hunt for Red October	Action Adventure
The Terminator	Action Sci-Fi





Summarizing Data with Functions



Summary Functions Available to SQL

Summary Function	Description
AVG, MEAN	Average or mean of values
COUNT, FREQ, N	Aggregate number of non-missing values
CSS	Corrected sum of squares
CV	Coefficient of variation
MAX	Largest value
MIN	Smallest value
NMISS	Number of non-missing values
PRT	Probability of a greater absolute value of Student's t
RANGE	Difference between the largest and smallest values
STD	Standard deviation
STDERR	Standard error of the mean
SUM	Sum of values
SUMWGT	Sum of the weight variable values which is 1
T	Testing the hypothesis that the population mean is zero
USS	Uncorrected sum of squares
VAR	Variance

Exercise #6

1. Click the Open folder  icon located at the top of the SAS Display Manager.
2. Select the SAS program named “Exer06” from the list.
3. The code should display in the SAS Editor.
4. Run the PROC SQL code by clicking the Submit  icon.
5. Let’s discuss the exercise and corresponding output.



Exercise #6 (continued)

The SQL procedure can be used to produce a single aggregate value by summarizing data down rows. Using the AVG function produces the average movie length for all “Comedy” movies.

```
PROC SQL;
```



```
SELECT AVG(LENGTH) AS Avg_Movie_Length
```

```
FROM SASUSER.MOVIES
```

```
WHERE UPCASE(CATEGORY) IN (“COMEDY”);
```

```
QUIT;
```

Exercise #7

1. Click the Open folder  icon located at the top of the SAS Display Manager.
2. Select the SAS program named “Exer07” from the list.
3. The code should display in the SAS Editor.
4. Run the PROC SQL code by clicking the Submit  icon.
5. Let’s discuss the exercise and corresponding output.



Exercise #7 (continued)

SQL can be used to summarize data across columns. Using the MAX function, the longest movie length is determined and then the difference between the maximum and current movie length.

```
PROC SQL;
```

```
SELECT MAX(LENGTH) AS Longest_Movie,  
       MAX(LENGTH) - LENGTH
```



```
AS Difference
```

```
FROM SASUSER.MOVIES;
```

```
QUIT;
```

<u>Longest_Movie</u>	<u>Difference</u>
194	17
194	91
194	97
194	78
194	64
...	...
194	76
194	70
194	59
194	86
194	93
194	0

Exercise #8

1. Click the Open folder  icon located at the top of the SAS Display Manager.
2. Select the SAS program named “Exer08” from the list.
3. The code should display in the SAS Editor.
4. Run the PROC SQL code by clicking the Submit  icon.
5. Let’s discuss the exercise and corresponding output.



Exercise #8 (continued)

To avoid having to compute the MAX function a second time, the CALCULATED keyword is used to determine the difference between the maximum and current movie length.

```
PROC SQL;
```

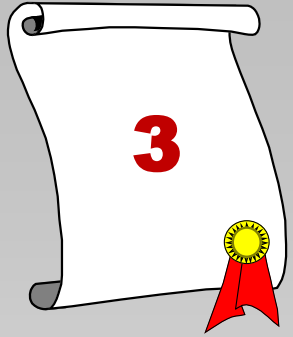
```
SELECT MAX(LENGTH) AS Longest_Movie,  
       CALCULATED Longest_Movie - LENGTH
```

```
       AS Difference
```

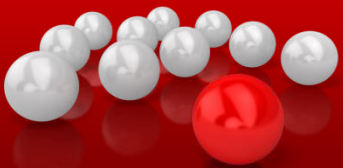
```
FROM SASUSER.MOVIES;
```

```
QUIT;
```

<u>Longest_Movie</u>	<u>Difference</u>
194	17
194	91
194	97
194	78
194	64
...	...
194	76
194	70
194	59
194	86
194	93
194	0



Exploring Data Relationships with Joins



What is a Join?

- Process of combining tables side-by-side
- Performs a matching process between rows in tables
- Some or all of the tables' contents are brought together
- Gather and manipulate data from across tables


Visually, it looks something like this:



Why Join?

- Data is often stored in separate tables
- Joins allow data to be combined as if it were stored in one huge file
- Provide exciting insights into data relationships
- Types of joins:
 - ✓ Inner joins – a maximum of 256 tables can be joined
 - ✓ Outer joins – a maximum of 2 tables can be joined

DATA Step Merge versus a Join

- A DATA step merge processes data differently than a standard join
 - The merge process overlays the duplicate by-column
 - PROC SQL adheres to ANSI (American National Standards Institute) guidelines
 - The join process doesn't overlay the duplicate matching column
- 
- A solid red bar with a diagonal cut across the bottom of the slide, starting from the left edge and extending towards the right edge.

Merge versus Join Results

Merge Results

Features

1. Data must be sorted by by-value.
2. More steps are needed than with SQL.
3. Requires common variable name.
4. Results are not automatically printed.

Join Results

Features

1. Duplicate matching column is not automatically overlaid as it is with merge.
2. Results are automatically printed unless NOPRINT option is specified.

What Happens During a Join?



When joining two tables:

- An intermediate Cartesian product is constructed from the two tables
- Rows are selected that match the WHERE clause

When joining more than two table:

- Determines the order of processing to reduce the size of the intermediate Cartesian product
- Reconstructs the join into several two-way joins
- Unwanted rows and columns from the intermediate tables are removed

Exercise #9

1. Click the Open folder  icon located at the top of the SAS Display Manager.
2. Select the SAS program named “Exer09” from the list.
3. The code should display in the SAS Editor.
4. Run the PROC SQL code by clicking the Submit  icon.
5. Let’s discuss the exercise and corresponding output.



Exercise #9 (continued)

Joining two tables together without a WHERE clause produces a Cartesian product consisting of a combination of all possible rows and columns.

```
PROC SQL;  
  SELECT *  
    FROM SASUSER.MOVIES,  
         SASUSER.ACTORS;  
QUIT;
```





Exercise #9 (continued)

The SAS System

Title	Length	Category	Year	Studio	Rating
Title	Actor	Leading	Actor	Supporting	
Brave Heart	177	Action Adventure	1995	Paramount Pictures	R
Brave Heart	Mel Gibson		Sophie Marceau		
Christmas Vacation	97	Comedy	1989	Warner Brothers	PG-13
Christmas Vacation	Chevy Chase		Beverly D'Angelo		
Coming to America	116	Comedy	1988	Paramount Pictures	R
Coming to America	Eddie Murphy		Arsenio Hall		
Forrest Gump	142	Drama	1994	Paramount Pictures	PG-13
Forrest Gump	Tom Hanks		Sally Field		
Ghost	127	Drama Romance	1990	Paramount Pictures	PG-13
Ghost	Patrick Swayze		Demi Moore		
Lethal Weapon	110	Action Cops & Robber	1987	Warner Brothers	R
Lethal Weapon	Mel Gibson		Danny Glover		

< Only Portion of Output Displayed >

Exercise #10

1. Click the Open folder  icon located at the top of the SAS Display Manager.
2. Select the SAS program named “Exer10” from the list.
3. The code should display in the SAS Editor.
4. Run the PROC SQL code by clicking the Submit  icon.
5. Let’s discuss the exercise and corresponding output.



Exercise #10 (continued)

The most reliable way to join two or more tables is with a WHERE clause. WHERE clauses are generally coded as a equi-join with an “=” sign.

```
PROC SQL;  
  SELECT MOVIES.title, rating, actor_leading  
    FROM SASUSER.MOVIES,  
         SASUSER.ACTORS  
   WHERE MOVIES.title = ACTORS.title;  
QUIT;
```





Exercise #10 (continued)

The SAS System

<u>Title</u>	<u>Rating</u>	<u>Actor Leading</u>
Brave Heart	R	Mel Gibson
Christmas Vacation	PG-13	Chevy Chase
Coming to America	R	Eddie Murphy
Forrest Gump	PG-13	Tom Hanks
Ghost	PG-13	Patrick Swayze
Lethal Weapon	R	Mel Gibson
Michael	PG-13	John Travolta
National Lampoon's Vacation	PG-13	Chevy Chase
Rocky	PG	Sylvester Stallone
Silence of the Lambs	R	Anthony Hopkins
The Hunt for Red October	PG	Sean Connery
The Terminator	R	Arnold Schwarzenegger
Titanic	PG-13	Leonardo DiCaprio

Exercise #11

1. Click the Open folder  icon located at the top of the SAS Display Manager.
2. Select the SAS program named “Exer11” from the list.
3. The code should display in the SAS Editor.
4. Run the PROC SQL code by clicking the Submit  icon.
5. Let’s discuss the exercise and corresponding output.



Exercise #11 (continued)

A left outer join with an ON clause matches rows from both tables and preserves all rows from the left table that did not have a match in the right table.

```
PROC SQL;  
  SELECT MOVIES.title, actor_leading, rating  
  FROM SASUSER.MOVIES  
  LEFT JOIN  
    SASUSER.ACTORS  
  ON MOVIES.title = ACTORS.title;  
QUIT;
```





Exercise #11 (continued)

The SAS System

Title	Actor Leading	Rating
Brave Heart	Mel Gibson	R
Casablanca		PG
Christmas Vacation	Chevy Chase	PG-13
Coming to America	Eddie Murphy	R
Dracula		R
Dressed to Kill		R
Forrest Gump	Tom Hanks	PG-13
Ghost	Patrick Swayze	PG-13
Jaws		PG
Jurassic Park		PG-13
Lethal Weapon	Mel Gibson	R
Michael	John Travolta	PG-13
National Lampoon's Vacation	Chevy Chase	PG-13
Poltergeist		PG
Rocky	Sylvester Stallone	PG
Scarface		R
Silence of the Lambs	Anthony Hopkins	R
Star Wars		PG
The Hunt for Red October	Sean Connery	PG
The Terminator	Arnold Schwarzenegge	R
The Wizard of Oz		G
Titanic	Leonardo DiCaprio	PG-13

Exercise #12

1. Click the Open folder  icon located at the top of the SAS Display Manager.
2. Select the SAS program named “Exer12” from the list.
3. The code should display in the SAS Editor.
4. Run the PROC SQL code by clicking the Submit  icon.
5. Let’s discuss the exercise and corresponding output.



Exercise #12 (continued)

A right outer join with an ON clause matches rows from both tables and preserves all rows from the right table that did not have a match in the left table.

```
PROC SQL;  
  SELECT MOVIES.title, actor_leading, rating  
  FROM SASUSER.MOVIES  
  RIGHT JOIN  
    SASUSER.ACTORS  
  ON MOVIES.title = ACTORS.title;  
QUIT;
```



Exercise #12 (continued)

The SAS System

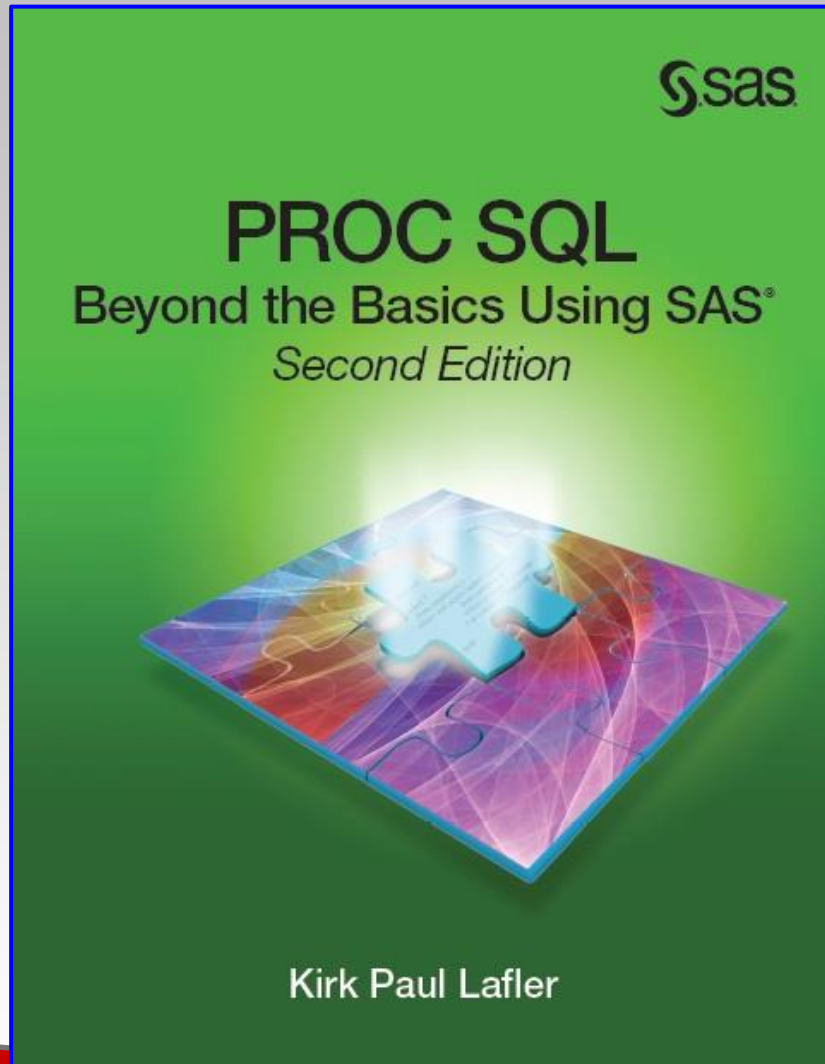
<u>Title</u>	<u>Actor Leading</u>	<u>Rating</u>
Brave Heart	Mel Gibson	R
Christmas Vacation	Chevy Chase	PG-13
Coming to America	Eddie Murphy	R
Forrest Gump	Tom Hanks	PG-13
Ghost	Patrick Swayze	PG-13
Lethal Weapon	Mel Gibson	R
Michael	John Travolta	PG-13
National Lampoon's Vacation	Chevy Chase	PG-13
Rocky	Sylvester Stallone	PG
Silence of the Lambs	Anthony Hopkins	R
The Hunt for Red October	Sean Connery	PG
The Terminator	Arnold Schwarzenegger	R
Titanic	Leonardo DiCaprio	PG-13

Conclusion

We've explored the following PROC SQL features for producing quick results to:

- **Construct simple queries to retrieve, subset, and search**
- **Summarize Data with Functions**
- **Explore data relationships with inner and outer joins**





**An SQL Book
with “under the
hood” details,
explanations and
lots of examples**

Available from SAS Press!

1st Edition



**Tips, Tricks and Shortcuts for
Better Searches and Better Results**

**Kirk Paul Lafler
Charles Edwin Shipp**

Odyssey
Press



**Become a Google
Search Pro!
Filled with
Tips, Tricks and Shortcuts
for Better Searches
and Better Results**

Available on www.Amazon.com!

Thank you for attending!

Quest  **ions?**

SAS® **GLOBAL FORUM** 2017

April 2 - 5 | Orlando, FL

A Hands-on Workshop by

Kirk Paul Lafler

KirkLafler@cs.com

@sasNerd