

Code Smarter, Not Harder: The 5 C's of ChatGPT for the SASsy Professional

Charu Shankar, SAS Institute;
Kirk Paul Lafler, sasNerd

ABSTRACT

In today's fast-paced analytics world, efficiency is everything. This paper explores how ChatGPT and SAS complement each other across the 5 C's to enhance productivity:

Communicate – Automate structured email reports with SAS, or use ChatGPT for dynamic, polished messaging. Learn when to use each for maximum impact.

Code – Speed up SAS programming with ChatGPT's syntax suggestions, debugging tips, and code generation—all while maintaining best practices.

Collaborate – Use SAS for version control and shared repositories, while ChatGPT streamlines teamwork with summaries, documentation, and code explanations.

Customize – Enhance efficiency with SAS macros and fine-tune ChatGPT prompts for tailored reporting and automation.

Create – Harness AI-driven insights for problem-solving while ensuring accuracy and compliance in SAS analytics.

Featuring hands-on demos and real-world examples, this session will equip you with practical strategies to code smarter, not harder.

INTRODUCTION

As data professionals, we're constantly seeking ways to work faster, communicate clearer, and deliver sharper insights. In an era defined by rapid decision-making and high data demands, efficiency isn't just nice to have—it's mission-critical. Enter the powerful duo of ChatGPT and SAS. Each tool excels on its own, but when combined strategically, they can transform the way we code, collaborate, and communicate.

This presentation introduces a practical framework based on the "5 C's"—Communicate, Code, Collaborate, Customize, and Create—that illustrates how to maximize productivity using the strengths of both platforms. Whether you're automating reports, debugging code, or summarizing results for your team, you'll learn when and how to leverage ChatGPT and SAS to work smarter—not harder.

DATASETS USED IN EXAMPLES

The examples presented in this paper illustrate two datasets, Movies and Actors. The Movies dataset, illustrated in Figure 1a, has 22 observations, a data structure with six variables, Title, Category, Studio, and Rating are defined as character variables; and Length and Year defined as numeric variables.

The Actors dataset, illustrated in Figure 1b, contains 13 observations and has a data structure consisting of three character-defined variables, Title, Actor_Leading and Actor_Supporting.

MOVIES Dataset

	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

Figure 1a: Movies dataset.

ACTORS Dataset

	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

Figure 1b: Actors dataset.

THE 5 C'S OF CHATGPT

COMMUNICATE

As data professionals, we're constantly seeking ways to work faster, communicate clearer, and deliver sharper insights. In an era defined by rapid decision-making and high data demands, efficiency isn't just nice to have—it's mission-critical. Enter the powerful duo of ChatGPT and SAS. Each tool excels on its own, but when combined strategically, they can transform the way we code, collaborate, and communicate.

ChatGPT provides amazing tools to improve writing, speaking, and conversational skills by crafting clear and succinct prose. A few productivity areas where ChatGPT excels includes the ability to:

- Generate initial drafts with clarity and conciseness.
- Improve grammatical accuracy by identifying and correcting errors.
- Practice conversational skills by providing feedback on pronunciation.

Communicate

Automate & Elevate Your Messaging

- **SAS** Auto-generate structured email reports
- **ChatGPT**: Craft dynamic, polished messages in seconds

Use Case:

- SAS builds the report → ChatGPT refines the summary for stakeholders

Pro Tip: Use SAS to deliver, ChatGPT to delight.

Copyright © SAS Institute Inc. All rights reserved.



```
proc ds2;  
  data _null_;  
    method init();  
    dcl varchar(16) str;  
    str = 'Hello World!';  
    put str;  
  end;  
  enddata;  
run;  
quit;
```

Program 1. Caption for Sample Program

Figure 1 is a sample figure.

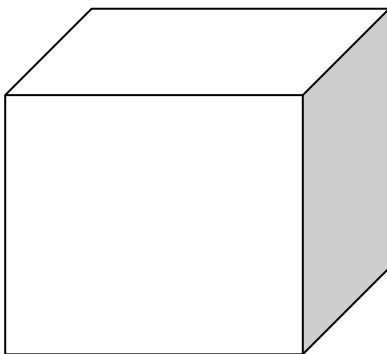


Figure 1. Caption for Sample Figure

CODE

Every coder runs into roadblocks—syntax slips, forgotten options, or logic bugs. ChatGPT acts like a virtual pair programmer, offering real-time suggestions, debugging help, and even full code examples. It can also help with understanding concepts and code optimization. While SAS remains the execution engine for reliable analysis, ChatGPT accelerates problem-solving and fosters a deeper understanding of SAS syntax and structure.

Code


Smarter SAS Programming Starts Here

- Syntax suggestions
- Debugging tips
- Generate boilerplate code

Use Case Stuck on PROC SQL? Ask ChatGPT for a template that works.

Pro Tip ChatGPT helps with quick answers, while **you maintain SAS best practices**

Copyright © SAS Institute Inc. All rights reserved.



COLLABORATE

Analytics projects rarely happen in isolation. Teams need tools that help them work better together. SAS offers powerful collaboration features like version control and audit trails, while ChatGPT fills the communication gaps—explaining code logic, translating outputs into lay terms, and quickly documenting complex workflows. The result is smoother teamwork and more inclusive data conversations.

Collaborate


Better Together: Teamwork, Streamlined

- **SAS** Version control & shared repositories
- **ChatGPT**: Summarize meetings, explain code, write documentation

Use Case ChatGPT explains that cryptic macro your coworker built in 2019.

Bridge technical and nontechnical team members with ~~an~~ powered clarity.

Copyright © SAS Institute Inc. All rights reserved.



CUSTOMIZE

Customization is key to efficiency. SAS macros provide repeatable automation for complex analytical tasks, making it easy to reuse logic and scale reporting. ChatGPT complements this with prompt engineering—generating summaries, cleaning narratives, and translating outputs into personalized messages. This blend of automation and AI responsiveness lets analysts serve different stakeholders without rewriting the wheel.

Customize

Tailor-Made Efficiency

- **SAS Macros** Automate recurring tasks with reusable logic
- **ChatGPT Prompts** Refine for personalized summaries, emails, code, and more

Use Case:

- SAS macro generates a report
- ChatGPT formats it in plain English (or executive-speak!)

Power Combo Macro the task, prompt the polish.

Copyright © SAS Institute Inc. All rights reserved.

CREATE

True innovation happens at the intersection of creativity and validation. ChatGPT encourages exploration by offering brainstorming support, hypothetical framing, and rapid prototyping of questions. SAS then steps in to test those ideas with rigorous, compliant statistical methods. This pairing not only enhances ideation but ensures the insights derived are defensible and actionable.

Create

From Insight to Impact

- Use ChatGPT to brainstorm, troubleshoot, and explore "what if" scenarios
- Validate and scale findings using SAS for trusted, compliant analytics

Use Case:

- ChatGPT helps frame the question
- SAS delivers data-driven answers with statistical rigor

AI + SAS = Insight that's innovative and accurate.

Copyright © SAS Institute Inc. All rights reserved.

CHATGPT CODE EXAMPLES

ChatGPT's code generation capabilities offer programmers and developers incredible opportunities for faster code development, debugging assistance, and enhanced productivity. We'll explore how ChatGPT can be used to translate natural language prompts into SAS and SQL code; generate code snippets; and how it can be used to optimize code for improved performance. To illustrate the power of ChatGPT with its prompts and code generation capabilities, the following examples will be introduced.

- SAS Code to Perform Exploratory Data Analysis (EDA)
- SAS Code to Create an Excel Spreadsheet
- SAS Code to Merge or Join Two Datasets
- SAS Code to Create Data-driven Alerts or Emails
- SAS Code to Perform Predictive Analytics

Produce SAS Code to Perform Exploratory Data Analysis (EDA)

Here's a SAS code template that was produced by ChatGPT to perform **Exploratory Data Analysis (EDA)** and includes:

- Descriptive statistics
- Frequency tables
- Histograms
- Box plots
- Correlation analysis (if numeric variables are present)
- Missing data checks

```
/* STEP 1: VIEW DATASET STRUCTURE */
PROC CONTENTS DATA=YOUR_DATASET;
RUN;

/* STEP 2: BASIC DESCRIPTIVE STATISTICS FOR NUMERIC VARIABLES */
PROC MEANS DATA=YOUR_DATASET N MEAN STD MIN MAX;
RUN;

/* STEP 3: FREQUENCY TABLES FOR CATEGORICAL VARIABLES */
PROC FREQ DATA=YOUR_DATASET;
  TABLES _CHARACTER_ / NOCUM NOPERCENT;
RUN;

/* STEP 4: CHECK FOR MISSING VALUES */
PROC FREQ DATA=YOUR_DATASET;
  TABLES _ALL_ / MISSING;
RUN;

/* STEP 5: VISUALIZE NUMERIC VARIABLES WITH HISTOGRAMS */
PROC UNIVARIATE DATA=YOUR_DATASET;
  VAR _NUMERIC_;
  HISTOGRAM;
RUN;

/* STEP 6: BOX PLOTS FOR NUMERIC VARIABLES GROUPED BY A CATEGORICAL VARIABLE */
PROC SGPLOT DATA=YOUR_DATASET;
  VBOX NUMERIC_VAR / CATEGORY=CATEGORICAL_VAR;
RUN;

/* STEP 7: CORRELATION ANALYSIS FOR NUMERIC VARIABLES */
PROC CORR DATA=YOUR_DATASET PLOTS=MATRIX(HISTOGRAM);
  VAR _NUMERIC_;
RUN;

/* STEP 8: SCATTERPLOT MATRIX FOR NUMERIC VARIABLES */
PROC SGSCATTER DATA=YOUR_DATASET;
  MATRIX _NUMERIC_;
RUN;
```

Produce SAS Code to Perform Exploratory Data Analysis (EDA), continued

From the SAS code template that ChatGPT produced, we'll modify it so that we can perform exploratory data analysis (EDA) on the MOVIES and ACTORS SAS datasets.

```
/* USE CHATGPT TO PRODUCE SAS CODE TO PERFORM EXPLORATORY DATA ANALYSIS (EDA) */

/* STEP 1: ACCESS SAS DATASET LIBRARY IN THE CLOUD */
LIBNAME MYDATA '/HOME/KIRKLAFLER/DATA SOURCES' ;

/* STEP 2: VIEW DATASET STRUCTURES */
PROC CONTENTS DATA=MYDATA.MOVIES;
RUN;
PROC CONTENTS DATA=MYDATA.ACTORS;
RUN;

/* STEP 3: BASIC DESCRIPTIVE STATISTICS FOR NUMERIC VARIABLES */
PROC MEANS DATA=MYDATA.MOVIES N NMISS MIN MAX RANGE MEAN MEDIAN MODE STD VAR;
RUN;

/* STEP 4: FREQUENCY TABLES FOR CATEGORICAL VARIABLES */
PROC FREQ DATA=MYDATA.MOVIES NLEVELS;
    TABLES _CHARACTER_ / NOCUM NOPERCENT;
RUN;
PROC FREQ DATA=MYDATA.ACTORS NLEVELS;
    TABLES _CHARACTER_ / NOCUM NOPERCENT;
RUN;

/* STEP 5: CHECK FOR MISSING VALUES */
PROC FREQ DATA=MYDATA.MOVIES;
    TABLES _ALL_ / MISSING;
RUN;
PROC FREQ DATA=MYDATA.ACTORS;
    TABLES _ALL_ / MISSING;
RUN;

/* STEP 6: VISUALIZE NUMERIC VARIABLES WITH HISTOGRAMS */
PROC UNIVARIATE DATA=MYDATA.MOVIES;
    VAR _NUMERIC_;
    HISTOGRAM;
RUN;

/* STEP 7: BOX PLOTS FOR NUMERIC VARIABLES GROUPED BY A CATEGORICAL VARIABLE */
PROC SGPLOT DATA=MYDATA.MOVIES;
    VBOX LENGTH / CATEGORY=RATING;
RUN;

/* STEP 8: CORRELATION ANALYSIS FOR NUMERIC VARIABLES */
PROC CORR DATA=MYDATA.MOVIES PLOTS=MATRIX(HISTOGRAM);
    VAR _NUMERIC_;
RUN;
```


Produce SAS Code to Perform Exploratory Data Analysis (EDA), continued

```
/* STEP 9: SCATTERPLOT MATRIX FOR NUMERIC VARIABLES */
PROC SGSCATTER DATA=MYDATA.MOVIES;
    MATRIX _NUMERIC_;
RUN;
```

Results:

The CONTENTS Procedure

Data Set Name	MYDATA.MOVIES	Observations	22
Member Type	DATA	Variables	6
Engine	V9	Indexes	0
Created	01/19/2025 18:42:01	Observation Length	128
Last Modified	01/19/2025 18:42:01	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information

Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	1022
Obs in First Data Page	22
Number of Data Set Repairs	0
Filename	/home/kirklafler/Data Sources/movies.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	11086108518
Access Permission	rw-r--r--
Owner Name	kirklafler
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
3	Category	Char	20	\$20.	\$20.
2	Length	Num	8	BEST12.	BEST32.
6	Rating	Char	5	\$5.	\$5.
5	Studio	Char	18	\$18.	\$18.
1	Title	Char	64	\$64.	\$64.
4	Year	Num	8	BEST12.	BEST32.

EDA Results, continued:

The CONTENTS Procedure

Data Set Name	MYDATA.ACTORS	Observations	13
Member Type	DATA	Variables	3
Engine	V9	Indexes	0
Created	11/22/2024 15:49:01	Observation Length	92
Last Modified	11/22/2024 15:49:01	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information

Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	1422
Obs in First Data Page	13
Number of Data Set Repairs	0
Filename	/home/kirklafler/Data Sources/actors.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	11086108517
Access Permission	rw-r--r--
Owner Name	kirklafler
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
2	Actor_Leading	Char	20	\$20.	\$20.
3	Actor_Supporting	Char	16	\$16.	\$16.
1	Title	Char	56	\$56.	\$56.

The MEANS Procedure

Variable	N	N Miss	Minimum	Maximum	Range	Mean	Median	Mode	Std Dev	Variance
Length	22	0	97.0000000	194.0000000	97.0000000	124.9090909	119.0000000	127.0000000	25.8344714	667.4199134
Year	22	0	1939.00	1997.00	58.0000000	1982.91	1987.50	1983.00	15.2124920	231.4199134

EDA Results, continued:

The FREQ Procedure

Number of Variable Levels	
Variable	Levels
Title	22
Category	11
Studio	10
Rating	4

Title	Frequency
Brave Heart	1
Casablanca	1
Christmas Vacation	1
Coming to America	1
Dracula	1
Dressed to Kill	1
Forrest Gump	1
Ghost	1
Jaws	1
Jurassic Park	1
Lethal Weapon	1
Michael	1
National Lampoon's Vacation	1
Poltergeist	1
Rocky	1
Scarface	1
Silence of the Lambs	1
Star Wars	1
The Hunt for Red October	1
The Terminator	1
The Wizard of Oz	1
Titanic	1

Category	Frequency
Action	1
Action Adventure	4
Action Cops & Robber	2
Action Sci-Fi	2
Adventure	1
Comedy	3
Drama	3
Drama Mysteries	1
Drama Romance	2
Drama Suspense	1
Horror	2

Studio	Frequency
Columbia TriStar	1
Filmways Pictures	1
Live Entertainment	1
Lucas Film Ltd	1
MGM / UA	4
Orion	1
Paramount Pictures	6
Universal Pictures	1
Universal Studios	2
Warner Brothers	4

Rating	Frequency
G	1
PG	6
PG13	7
R	8

EDA Results, continued:

The FREQ Procedure

Number of Variable Levels	
Variable	Levels
Title	13
Actor_Leading	11
Actor_Supporting	12

Title	Frequency
Brave Heart	1
Christmas Vacation	1
Coming to America	1
Forrest Gump	1
Ghost	1
Lethal Weapon	1
Michael	1
National Lampoon's Vacation	1
Rocky	1
Silence of the Lambs	1
The Hunt for Red October	1
The Terminator	1
Titanic	1

Actor_Leading	Frequency
Anthony Hopkins	1
Arnold Schwarzenegger	1
Chevy Chase	2
Eddie Murphy	1
John Travolta	1
Leonardo DiCaprio	1
Mel Gibson	2
Patrick Swayze	1
Sean Connery	1
Sylvester Stallone	1
Tom Hanks	1

Actor_Supporting	Frequency
Alec Baldwin	1
Andie MacDowell	1
Arsenio Hall	1
Beverly D'Angelo	2
Danny Glover	1
Demi Moore	1
Jodie Foster	1
Kate Winslet	1
Michael Biehn	1
Sally Field	1
Sophie Marceau	1
Talia Shire	1

The UNIVARIATE Procedure Variable: Length

Moments			
N	22	Sum Weights	22
Mean	124.909091	Sum Observations	2748
Std Deviation	25.8344714	Variance	667.419913
Skewness	1.45414494	Kurtosis	1.71453814
Uncorrected SS	357266	Corrected SS	14015.8182
Coeff Variation	20.682619	Std Error Mean	5.50792781

Basic Statistical Measures			
Location		Variability	
Mean	124.9091	Std Deviation	25.83447
Median	119.0000	Variance	667.41991
Mode	127.0000	Range	97.00000
		Interquartile Range	24.00000

Tests for Location: Mu0=0				
Test	Statistic	p Value		
Student's t	t	22.67806	Pr > t	<.0001
Sign	M	11	Pr >= M	<.0001
Signed Rank	S	126.5	Pr >= S	<.0001

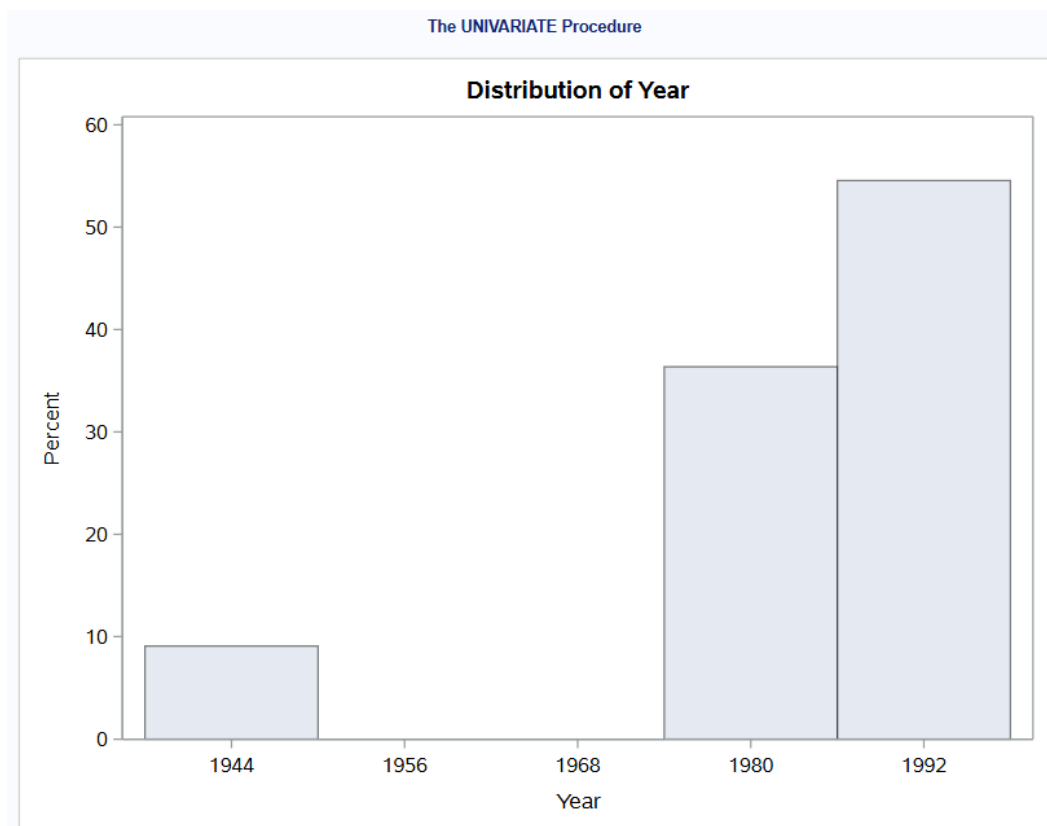
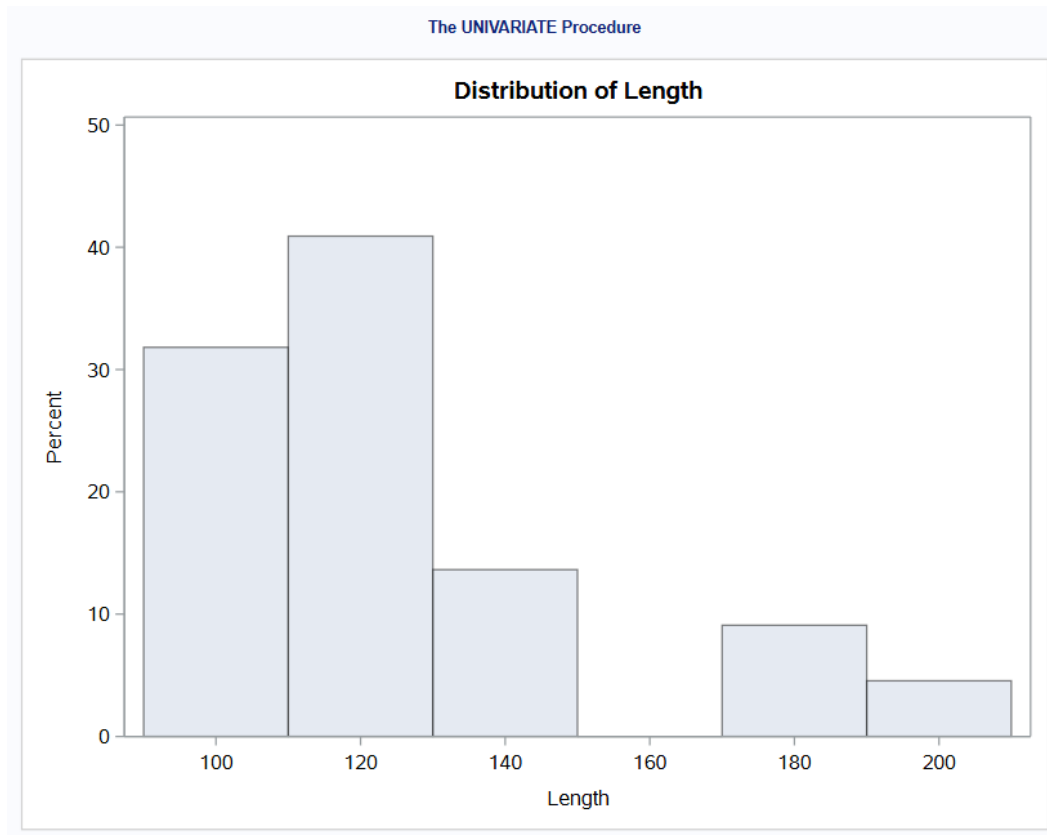
Quantiles (Definition 5)

Level	Quantile
100% Max	194
99%	194
95%	177
90%	170
75% Q3	130
50% Median	119
25% Q1	106
10%	101
5%	98
1%	97
0% Min	97

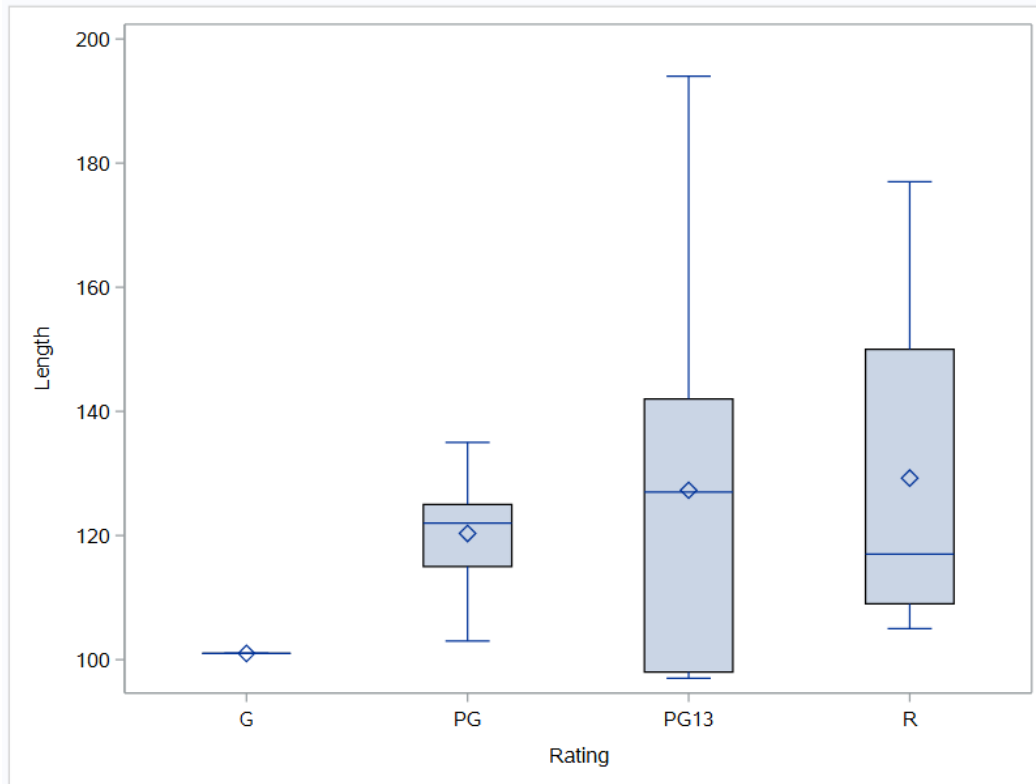
Extreme Observations

Lowest		Highest	
Value	Obs	Value	Obs
97	3	135	19
98	13	142	7
101	21	170	16
103	2	177	1
105	6	194	22

EDA Results, continued:



EDA Results, continued:



The CORR Procedure

2 Variables: Length Year

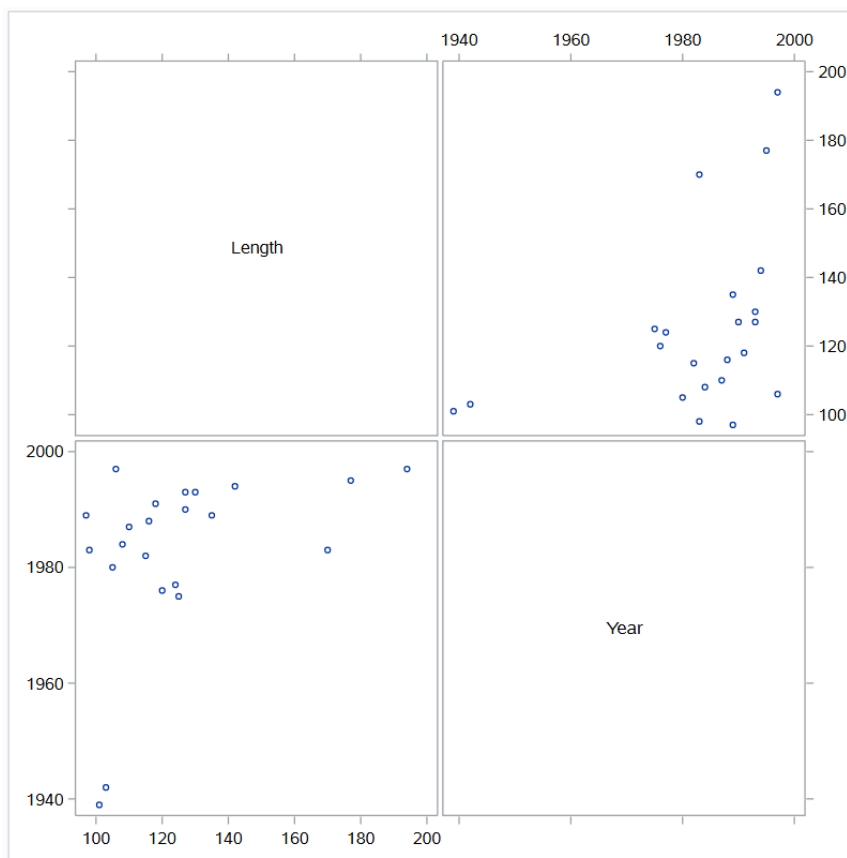
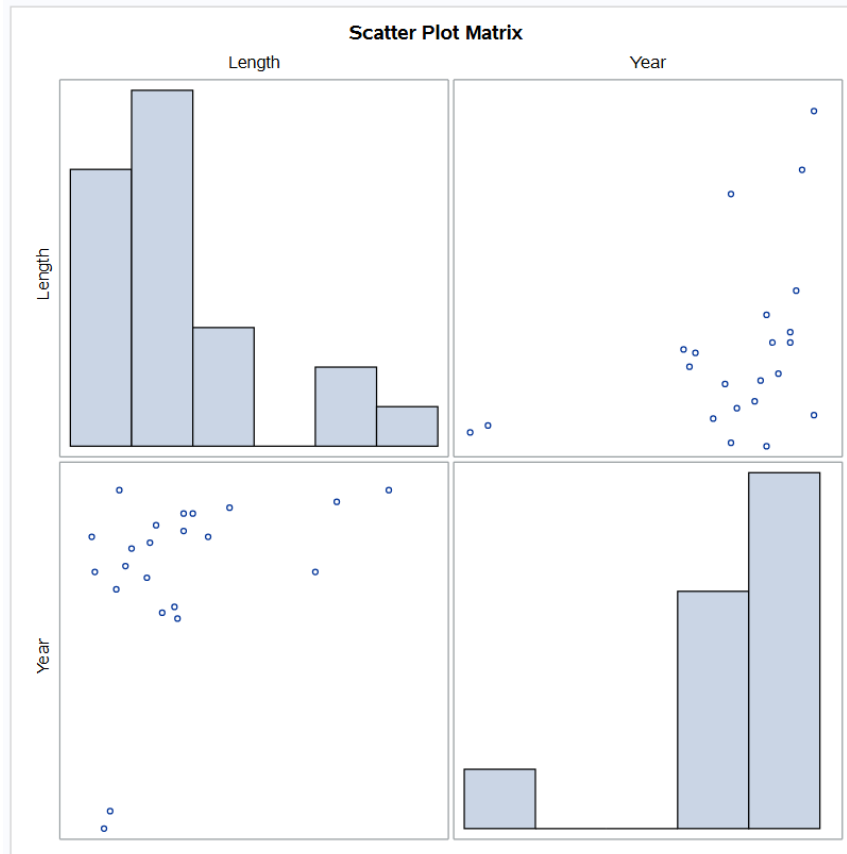
Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Length	22	124.90909	25.83447	2748	97.00000	194.00000
Year	22	1983	15.21249	43624	1939	1997

Pearson Correlation Coefficients, N = 22 Prob > |r| under H0: Rho=0

	Length	Year
Length	1.00000	0.40928 0.0586
Year	0.40928 0.0586	1.00000

EDA Results, continued:



Produce SAS Code to Create an Excel Spreadsheet

Here's an example template of **SAS code** to create an **Excel spreadsheet** using the ODS EXCEL statement, which is the easiest and most flexible way in modern SAS (especially SAS 9.4+).

```
/* SET THE OUTPUT DESTINATION TO AN EXCEL FILE */
ODS EXCEL FILE="C:\PATH\TO\YOUR\OUTPUT.XLSX"
      OPTIONS (SHEET_NAME="MYDATASHEET") ;

/* CREATE SOME SAMPLE DATA */
DATA MYDATA;
  INPUT NAME $ AGE HEIGHT WEIGHT;
  DATALINES;
JOHN 25 68 150
SARA 30 65 120
MIKE 22 70 175
;
RUN;

/* PRODUCE A REPORT TO THE EXCEL FILE */
PROC REPORT DATA=MYDATA;
  COLUMNS VARIABLES-LIST;
  DEFINE ATTRIBUTES ASSOCIATED WITH VARIABLE-LIST;
RUN;

/* CLOSE THE ODS EXCEL DESTINATION */
ODS EXCEL CLOSE;
```

Key Points:

- ods excel file=... tells SAS to start writing to an Excel file.
- sheet_name= lets you name the Excel sheet.
- You can use any PROC (like PROC PRINT, PROC REPORT, PROC TABULATE, etc.) between ods excel and ods excel close.
- After all procedures, you must always end with ods excel close; to properly save the file.

Produce SAS Code to Create an Excel Spreadsheet, continued

From the SAS code template that ChatGPT produced, we'll modify it so that we can create an Excel spreadsheet using the MOVIES SAS dataset.

```
/* USE CHATGPT TO PRODUCE SAS CODE TO CREATE AN EXCEL SPREADSHEET */

/* STEP 1: ACCESS SAS DATASET LIBRARY IN THE CLOUD */
LIBNAME MYDATA '/HOME/KIRKLAFLER/DATA SOURCES' ;

/* SET THE OUTPUT DESTINATION TO AN EXCEL FILE */
ODS EXCEL FILE="/HOME/KIRKLAFLER/CHATGPT AND SAS/RESULTS/CHATGPT-TO-SAS-EXCEL.XLSX"
    OPTIONS(SHEET_NAME="MY-CHATGPT-DATASHEET") ;

/* PRODUCE REPORT RESULTS TO AN EXCEL FILE */
PROC REPORT DATA=MYDATA.MOVIES;
    COLUMNS CATEGORY TITLE LENGTH STUDIO YEAR RATING;
    DEFINE CATEGORY / ORDER;
    DEFINE TITLE / DISPLAY;
    DEFINE LENGTH / DISPLAY;
    DEFINE STUDIO / DISPLAY;
    DEFINE YEAR / DISPLAY;
    DEFINE RATING / DISPLAY;
RUN;

/* CLOSE THE ODS EXCEL DESTINATION */
ODS EXCEL CLOSE;
```

Results:

1	Category	Title	Length	Studio	Year	Rating
2	Action	Jurassic Park	127	Universal Pictures	1993	PG13
3	Action Adventure	Brave Heart	177	Paramount Pictures	1995	R
4		Jaws	125	Universal Studios	1975	PG
5		Rocky	120	MGM / UA	1976	PG
6		The Hunt for Red October	135	Paramount Pictures	1989	PG
7	Action Cops & Robber	Lethal Weapon	110	Warner Brothers	1987	R
8		Scarface	170	Universal Studios	1983	R
9	Action Sci-Fi	Star Wars	124	Lucas Film Ltd	1977	PG
10		The Terminator	108	Live Entertainment	1984	R
11	Adventure	The Wizard of Oz	101	MGM / UA	1939	G
12	Comedy	Christmas Vacation	97	Warner Brothers	1989	PG13
13		Coming to America	116	Paramount Pictures	1988	R
14		National Lampoon's Vacation	98	Warner Brothers	1983	PG13
15	Drama	Casablanca	103	MGM / UA	1942	PG
16		Forrest Gump	142	Paramount Pictures	1994	PG13
17		Michael	106	Warner Brothers	1997	PG13
18	Drama Mysteries	Dressed to Kill	105	Filmways Pictures	1980	R
19	Drama Romance	Ghost	127	Paramount Pictures	1990	PG13
20		Titanic	194	Paramount Pictures	1997	PG13
21	Drama Suspense	Silence of the Lambs	118	Orion	1991	R
22	Horror	Dracula	130	Columbia TriStar	1993	R
23		Poltergeist	115	MGM / UA	1982	PG
24						
25						

< > MY-CHATGPT-DATASHEET +

Produce SAS Code to Merge or Join Two Datasets

ChatGPT produces a SAS code template to **merge (or join) two datasets in SAS** — both **basic merge** and **SQL join** styles.

METHOD 1: USING MERGE IN A DATA STEP (BY KEY)

✓ SORT AND MERGE

```
/* THIS IS A CLASSIC SORT AND MERGE USING A COMMON KEY VARIABLE.
   MAKE SURE BOTH DATASETS ARE SORTED BY THE COMMON_KEY BEFOREHAND.
   THE IN=A AND IN=B ARE OPTIONAL FLAGS TO TRACK SOURCE OF OBSERVATIONS. */
PROC SORT DATA=DATASET1 OUT=DATASET1_SORTED;
  BY COMMON_KEY;
RUN;
PROC SORT DATA=DATASET2 OUT=DATASET2_SORTED;
  BY COMMON_KEY;
RUN;
DATA MERGED_DATA;
  MERGE DATASET1_SORTED (IN=A)
        DATASET2_SORTED (IN=B);
  BY COMMON_KEY;
RUN;
```

METHOD 2: USING PROC SQL FOR SQL-STYLE JOINS

✓ INNER JOIN

```
PROC SQL;
  CREATE TABLE INNER_JOINED_DATA AS
  SELECT A.*, B.*
  FROM DATASET1 AS A
  INNER JOIN
    DATASET2 AS B
  ON A.COMMON_KEY = B.COMMON_KEY;
QUIT;
```

✓ LEFT JOIN

```
PROC SQL;
  CREATE TABLE LEFT_JOINED_DATA AS
  SELECT A.*, B.*
  FROM DATASET1 AS A
  LEFT JOIN
    DATASET2 AS B
  ON A.COMMON_KEY = B.COMMON_KEY;
QUIT;
```

Produce SAS Code to Merge or Join Two Datasets, continued

METHOD 1: USING MERGE IN A DATA STEP (BY KEY)

From the SAS code template that ChatGPT produced, we'll modify it so that we can merge the MOVIES and ACTORS SAS datasets and then produce metadata and physical data from the “merged” results.

```
/* THIS IS A CLASSIC SORT AND MERGE USING A COMMON KEY VARIABLE.
   MAKE SURE BOTH DATASETS ARE SORTED BY THE COMMON_KEY BEFOREHAND.
   THE IN=M AND IN=A ARE OPTIONAL FLAGS TO TRACK SOURCE OF OBSERVATIONS. */

/* STEP 1: ACCESS SAS DATASET LIBRARY IN THE CLOUD */
LIBNAME MYDATA '/HOME/KIRKLAFLER/DATA SOURCES' ;

PROC SORT DATA=MYDATA.MOVIES OUT=WORK.MOVIES_SORTED;
  BY TITLE;
RUN;

PROC SORT DATA=MYDATA.ACTORS OUT=WORK.ACTORS_SORTED;
  BY TITLE;
RUN;

DATA WORK.MERGED_DATA;
  MERGE WORK.MOVIES_SORTED (IN=M)
        WORK.ACTORS_SORTED (IN=A);
  BY TITLE;
RUN;

PROC CONTENTS DATA=WORK.MERGED_DATA;
RUN;

PROC PRINT DATA=WORK.MERGED_DATA;
RUN;
```

Results:

The CONTENTS Procedure			
Data Set Name	WORK.MERGED_DATA	Observations	22
Member Type	DATA	Variables	8
Engine	V9	Indexes	0
Created	05/02/2025 16:54:02	Observation Length	160
Last Modified	05/02/2025 16:54:02	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Produce SAS Code to Merge or Join Two Datasets, continued

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	818
Obs in First Data Page	22
Number of Data Set Repairs	0
Filename	/saswork/SAS_work2B0800002D67_odaws01-usw2.oda.sas.com/SAS_workA04B00002D67_odaws01-usw2.oda.sas.com/merged_data.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	3435041
Access Permission	rw-r--r--
Owner Name	kirklafler
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
7	Actor_Leading	Char	20	\$20.	\$20.
8	Actor_Supporting	Char	16	\$16.	\$16.
3	Category	Char	20	\$20.	\$20.
2	Length	Num	8	BEST12.	BEST32.
6	Rating	Char	5	\$5.	\$5.
5	Studio	Char	18	\$18.	\$18.
1	Title	Char	64	\$64.	\$64.
4	Year	Num	8	BEST12.	BEST32.

Obs	Title	Length	Category	Year	Studio	Rating	Actor_Leading	Actor_Supporting
1	Brave Heart	177	Action Adventure	1995	Paramount Pictures	R	Mel Gibson	Sophie Marceau
2	Casablanca	103	Drama	1942	MGM / UA	PG		
3	Christmas Vacation	97	Comedy	1989	Warner Brothers	PG13	Chevy Chase	Beverly D'Angelo
4	Coming to America	116	Comedy	1988	Paramount Pictures	R	Eddie Murphy	Arsenio Hall
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	PG13	Tom Hanks	Sally Field
8	Ghost	127	Drama Romance	1990	Paramount Pictures	PG13	Patrick Swayze	Demi Moore
9	Jaws	125	Action Adventure	1975	Universal Studios	PG		
10	Jurassic Park	127	Action	1993	Universal Pictures	PG13		
11	Lethal Weapon	110	Action Cops & Robber	1987	Warner Brothers	R	Mel Gibson	Danny Glover
12	Michael	106	Drama	1997	Warner Brothers	PG13	John Travolta	Andie MacDowell
13	National Lampoon's Vacation	98	Comedy	1983	Warner Brothers	PG13	Chevy Chase	Beverly D'Angelo
14	Poltergeist	115	Horror	1982	MGM / UA	PG		
15	Rocky	120	Action Adventure	1976	MGM / UA	PG	Sylvester Stallone	Talia Shire
16	Scarface	170	Action Cops & Robber	1983	Universal Studios	R		
17	Silence of the Lambs	118	Drama Suspense	1991	Orion	R	Anthony Hopkins	Jodie Foster
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	Sean Connery	Alec Baldwin
20	The Terminator	108	Action Sci-Fi	1984	Live Entertainment	R	Arnold Schwarzenegger	Michael Biehn
21	The Wizard of Oz	101	Adventure	1939	MGM / UA	G		
22	Titanic	194	Drama Romance	1997	Paramount Pictures	PG13	Leonardo DiCaprio	Kate Winslet

Produce SAS Code to Merge or Join Two Datasets, continued

METHOD 2: USING PROC SQL FOR SQL-STYLE JOINS

✓ INNER JOIN

```
/* THIS IS A CLASSIC SQL INNER JOIN QUERY USING PROC SQL AND A COMMON KEY VARIABLE. */

/* STEP 1: ACCESS SAS DATASET LIBRARY IN THE CLOUD */
LIBNAME MYDATA '/HOME/KIRKLAFLER/DATA SOURCES' ;

PROC SQL;
  CREATE TABLE WORK INNER_JOIN_DATA AS
  SELECT M.TITLE,
         M.LENGTH,
         M.CATEGORY,
         M.STUDIO,
         M.YEAR,
         M.RATING,
         A.ACTOR_LEADING,
         A.ACTOR_SUPPORTING
  FROM MYDATA.MOVIES AS M
       INNER JOIN
       MYDATA.ACTOR AS A
       ON M.TITLE = A.TITLE;
QUIT;

PROC CONTENTS DATA=WORK INNER_JOIN_DATA;
RUN;

PROC PRINT DATA=WORK INNER_JOIN_DATA;
RUN;
```

Results:

The CONTENTS Procedure

Data Set Name	WORK INNER_JOIN_DATA	Observations	13
Member Type	DATA	Variables	8
Engine	V9	Indexes	0
Created	05/02/2025 17:23:30	Observation Length	160
Last Modified	05/02/2025 17:23:30	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Produce SAS Code to Merge or Join Two Datasets, continued

Inner Join Results, continued:

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	818
Obs in First Data Page	13
Number of Data Set Repairs	0
Filename	/saswork/SAS_work2B0800002D67_odaws01-usw2.oda.sas.com/SAS_workA04B00002D67_odaws01-usw2.oda.sas.com/inner_join_data.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	3402697
Access Permission	rw-r--r--
Owner Name	kirklafler
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
7	Actor_Leading	Char	20	\$20.	\$20.
8	Actor_Supporting	Char	16	\$16.	\$16.
3	Category	Char	20	\$20.	\$20.
2	Length	Num	8	BEST12.	BEST32.
6	Rating	Char	5	\$5.	\$5.
4	Studio	Char	18	\$18.	\$18.
1	Title	Char	64	\$64.	\$64.
5	Year	Num	8	BEST12.	BEST32.

Obs	Title	Length	Category	Studio	Year	Rating	Actor_Leading	Actor_Supporting
1	Brave Heart	177	Action Adventure	Paramount Pictures	1995	R	Mel Gibson	Sophie Marceau
2	Christmas Vacation	97	Comedy	Warner Brothers	1989	PG13	Chevy Chase	Beverly D'Angelo
3	Coming to America	116	Comedy	Paramount Pictures	1988	R	Eddie Murphy	Arsenio Hall
4	Forrest Gump	142	Drama	Paramount Pictures	1994	PG13	Tom Hanks	Sally Field
5	Ghost	127	Drama Romance	Paramount Pictures	1990	PG13	Patrick Swayze	Demi Moore
6	Lethal Weapon	110	Action Cops & Robber	Warner Brothers	1987	R	Mel Gibson	Danny Glover
7	Michael	106	Drama	Warner Brothers	1997	PG13	John Travolta	Andie MacDowell
8	National Lampoon's Vacation	98	Comedy	Warner Brothers	1983	PG13	Chevy Chase	Beverly D'Angelo
9	Rocky	120	Action Adventure	MGM / UA	1976	PG	Sylvester Stallone	Talia Shire
10	Silence of the Lambs	118	Drama Suspense	Orion	1991	R	Anthony Hopkins	Jodie Foster
11	The Hunt for Red October	135	Action Adventure	Paramount Pictures	1989	PG	Sean Connery	Alec Baldwin
12	The Terminator	108	Action Sci-Fi	Live Entertainment	1984	R	Arnold Schwarzenegger	Michael Biehn
13	Titanic	194	Drama Romance	Paramount Pictures	1997	PG13	Leonardo DiCaprio	Kate Winslet

Produce SAS Code to Merge or Join Two Datasets, continued

✓ LEFT JOIN

```
PROC SQL;
  CREATE TABLE LEFT_JOINED_DATA AS
  SELECT A.*, B.*
  FROM DATASET1 AS A
  LEFT JOIN
    DATASET2 AS B
  ON A.COMMON_KEY = B.COMMON_KEY;
QUIT;
```

Results:

The CONTENTS Procedure

Data Set Name	WORK.LEFT_JOIN_DATA	Observations	22
Member Type	DATA	Variables	8
Engine	V9	Indexes	0
Created	05/02/2025 17:35:01	Observation Length	160
Last Modified	05/02/2025 17:35:01	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	YES
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information

Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	818
Obs in First Data Page	22
Number of Data Set Repairs	0
Filename	/saswork/SAS_work2B0800002D67_odaws01-usw2.oda.sas.com/SAS_workA04B00002D67_odaws01-usw2.oda.sas.com/left_join_data.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	3430463
Access Permission	rw-r--r--
Owner Name	kirklafler
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
7	Actor_Leading	Char	20	\$20.	\$20.
8	Actor_Supporting	Char	16	\$16.	\$16.
3	Category	Char	20	\$20.	\$20.
2	Length	Num	8	BEST12.	BEST32.
6	Rating	Char	5	\$5.	\$5.
4	Studio	Char	18	\$18.	\$18.
1	Title	Char	64	\$64.	\$64.
5	Year	Num	8	BEST12.	BEST32.

Produce SAS Code to Merge or Join Two Datasets, continued

Left Join Results, continued:

Obs	Title	Length	Category	Studio	Year	Rating	Actor_Leading	Actor_Supporting
1	Brave Heart	177	Action Adventure	Paramount Pictures	1995	R	Mel Gibson	Sophie Marceau
2	Casablanca	103	Drama	MGM / UA	1942	PG		
3	Christmas Vacation	97	Comedy	Warner Brothers	1989	PG13	Chevy Chase	Beverly D'Angelo
4	Coming to America	116	Comedy	Paramount Pictures	1988	R	Eddie Murphy	Arsenio Hall
5	Dracula	130	Horror	Columbia TriStar	1993	R		
6	Dressed to Kill	105	Drama Mysteries	Filmways Pictures	1980	R		
7	Forrest Gump	142	Drama	Paramount Pictures	1994	PG13	Tom Hanks	Sally Field
8	Ghost	127	Drama Romance	Paramount Pictures	1990	PG13	Patrick Swayze	Demi Moore
9	Jaws	125	Action Adventure	Universal Studios	1975	PG		
10	Jurassic Park	127	Action	Universal Pictures	1993	PG13		
11	Lethal Weapon	110	Action Cops & Robber	Warner Brothers	1987	R	Mel Gibson	Danny Glover
12	Michael	106	Drama	Warner Brothers	1997	PG13	John Travolta	Andie MacDowell
13	National Lampoon's Vacation	98	Comedy	Warner Brothers	1983	PG13	Chevy Chase	Beverly D'Angelo
14	Poltergeist	115	Horror	MGM / UA	1982	PG		
15	Rocky	120	Action Adventure	MGM / UA	1976	PG	Sylvester Stallone	Talia Shire
16	Scarface	170	Action Cops & Robber	Universal Studios	1983	R		
17	Silence of the Lambs	118	Drama Suspense	Orion	1991	R	Anthony Hopkins	Jodie Foster
18	Star Wars	124	Action Sci-Fi	Lucas Film Ltd	1977	PG		
19	The Hunt for Red October	135	Action Adventure	Paramount Pictures	1989	PG	Sean Connery	Alec Baldwin
20	The Terminator	108	Action Sci-Fi	Live Entertainment	1984	R	Arnold Schwarzenegger	Michael Biehn
21	The Wizard of Oz	101	Adventure	MGM / UA	1939	G		
22	Titanic	194	Drama Romance	Paramount Pictures	1997	PG13	Leonardo DiCaprio	Kate Winslet

Produce SAS Code to Create Data-driven Alerts or Emails

We'll now show you an SAS **example** to produce **data-driven alerts** (or **emails**) — meaning **SAS can automatically send an alert (or email)** when certain conditions in your data are met.

Here's a full example SAS program:

```
/* Step 1: Set up the email system (SMTP) */
options emailsys=smtp
      emailhost="smtp.yourserver.com" /* your SMTP server */
      emailport=587                  /* SMTP port */
      emailauthprotocol=LOGIN
      emailid="your_email@example.com"
      emailpw="your_password";      /* be careful with passwords */

/* Step 2: Check the data */
data alerts;
  set your_dataset; /* Your data source */
  /* Define the alert condition, e.g., sales too low */
  if sales < 1000 then do;
    call symputx('trigger_alert', 'YES');
    call symputx('alert_detail', cat('Sales dropped to ', sales));
  end;
run;

/* Step 3: Send an email if an alert is triggered */
%macro send_alert;
  %if &trigger_alert = YES %then %do;
    filename myemail email
      to="recipient@example.com"
      subject="ALERT: Sales Dropped!"
      from="your_email@example.com";
    data _null_;
      file myemail;
      put "Attention,";
      put "";
      put "An alert has been triggered:";
      put "&alert_detail.";
      put "";
      put "Please investigate immediately.";
    run;
  %end;
%mend send_alert;

%send_alert;
```

Quick overview of what's happening:

- It **reads your data** (your_dataset).
- It **checks for a condition** (example: sales < 1000).
- If triggered, it **sends an automated email**.

NOTES:

- You must have SMTP settings configured (options emailsys=SMTP) for it to work.
- You can make this more dynamic with loops, HTML formatting, attachments, etc.

Produce SAS Code to Create Data-driven Alerts or Emails, continued

From the SAS code template that ChatGPT produced, we'll modify it so that we can create a data-driven alert or email using the MOVIES SAS dataset.

```
/* THIS IS A CLASSIC TEMPLATE TO CREATE DATA-DRIVEN ALERTS OR EMAILS. */

/* Step 1: Access SAS Dataset Library in the Cloud */
LIBNAME MYDATA '/home/kirklafler/Data Sources' ;

/* Step 2: Check the data */
data alerts;
  set MYDATA.MOVIES; /* SAS data source */
  /* Define the alert condition, e.g., recent movie */
  if YEAR GE MAX(YEAR) then do;
    call symputx('trigger_alert', 'YES');
    call symputx('alert_title', cat ('Movie: ', TITLE));
    call symputx('alert_year', cat('Year: ', YEAR));
  end;
run;

/* Step 3: Send an email if an alert is triggered */
%macro send_alert;
  %if &trigger_alert = YES %then %do;
    filename mymail email
      to="KirkLafler@cs.com"
      subject="ALERT: Recent Movie(s) is/are Now Available!"
      from="KirkLafler@cs.com";
    data _null_;
      file mymail;
      put "Attention Movie Fans:";
      put ;
      put "An alert has been triggered!";
      put ;
      put "&alert_title., &alert_year. is available for purchase";
      put ;
      put "Orders will be processed within 3 days.";
    run;
  %end;
%mend send_alert;

%send_alert;
```

E-mail Alert:

• ALERT: Recent Movie(s) is/are Now Available!

Old Mail ☆



• odaadmin@sas.com
From: odaadmin@sas.com
To: kirklafler@cs.com



Fri, May 2 at 10:23 PM ☆

Attention Movie Fans,

An alert has been triggered:

Movie: Titanic, Year: 1997 is available for purchase

Orders will be processed within 3 days.

Produce SAS code to perform predictive analytics on a Movies dataset with the following variables: Title, Length, Category, Studio, Year, and Rating where Rating is character and contains the values: G, PG, PG-13, and R.

ChatGPT provides a sample SAS program that performs **predictive analytics** on the described Movies dataset. The goal here will be to **predict the movie rating** (Rating) using the other variables as predictors. Since Rating is a **categorical variable**, this is a **classification problem**, and a good starting point is to use **decision trees** (via PROC HPSPLIT) or **logistic regression (multinomial)**.

Assumptions:

- Rating is categorical (or character and contains values of "G", "PG", "PG-13", and "R").
- Category and Studio are categorical.
- Data is in a dataset called Movies.

```
/* Step 1: Data Access and Preparation */
libname mydata '/your/directory/path'; /* Set library path */

/* Convert categorical variables to class variables */
data work.movies_prep;
  set MYDATA.MOVIES;
  length RatingCat $10;
  RatingCat = strip(Rating); /* Ensure there's no leading/trailing space */
run;

/* Step 2: Multinomial Logistic Regression Using PROC LOGISTIC to predict Rating */
/* The link=glogit option is used for generalized logits (multinomial outcomes). */
/* You can change the reference category (ref='PG') as needed. */
proc logistic data=work.movies_prep;
  class Category Studio RatingCat (ref='PG') / param=glm;
  model RatingCat = Length Category Studio Year / link=glogit;
run;

/* Step 3: Decision Tree Classification Using PROC HPSPLIT */
/* Notes: grow entropy: uses information gain for splitting. */
/*        prune costcomplexity: helps reduce overfitting. */
/*        code generates scoring code to apply the model later. */
proc hpsplit data=work.movies_prep;
  class Category Studio RatingCat;
  model RatingCat = Length Category Studio Year;
  grow entropy;
  prune costcomplexity;
  code file='tree_score.sas';
run;

/* Step 4: Optional: Splitting Data Into Training and Validation */
proc surveyselect data=work.movies_prep
  out=work.movies_split
  samprate=0.7 outall seed=12345;
run;

data work.train work.valid;
  set work.movies_split;
  if selected=1 then output work.train;
  else output work.valid;
run;
```

Produce SAS Code to Perform Predictive Analytics, continued

```
/* Step 5: Visualize the Decision Tree */
proc hpsplit data=work.train seed=123;
  class Category Studio RatingCat;
  model RatingCat = Length Category Studio Year;
  grow entropy;
  prune costcomplexity;
  partition fraction(validate=0.3); /* Auto split for validation */
  code file='tree_score.sas'; /* For scoring later */
  output out=work.tree_out; /* For later analysis */
run;

/* Step 6: Step 2: Score the Validation Dataset */
data work.valid_scored;
  set work.valid;
  %include 'tree_score.sas'; /* Apply the decision tree model */
run;

/* Step 7: Evaluate Model Accuracy (Confusion Matrix) */
proc freq data=work.valid_scored;
  tables RatingCat*P_RatingCat / norow nocol nopercent;
  title 'Confusion Matrix: Actual vs Predicted Ratings';
run;

/* Step 8: ROC Curves (One-vs-Rest for Multiclass) */
/* Notes: ROC curves and variable importance plots will give more insight */
/*         into your model's performance and how each variable contributes. */
proc logistic data=work.train plots=roc;
  class Category Studio RatingCat (ref='PG') / param=glm;
  model RatingCat(event='G') = Length Category Studio Year;
  title 'ROC Curve: Predicting G Rating vs Others';
run;

proc logistic data=work.train plots=roc;
  class Category Studio RatingCat (ref='PG') / param=glm;
  model RatingCat(event='PG-13') = Length Category Studio Year;
  title 'ROC Curve: Predicting PG-13 Rating vs Others';
run;

proc logistic data=work.train plots=roc;
  class Category Studio RatingCat (ref='PG') / param=glm;
  model RatingCat(event='R') = Length Category Studio Year;
  title 'ROC Curve: Predicting R Rating vs Others';
run;

/* Step 9: Variable Importance (from Decision Tree) */
proc hpsplit data=work.train seed=123;
  class Category Studio RatingCat;
  model RatingCat = Length Category Studio Year;
  grow entropy;
  prune costcomplexity;
  partition fraction(validate=0.3);
  code file='tree_score.sas';
  output out=work.tree_out;
  importance;
  title 'Variable Importance from Decision Tree';
run;
```

Produce SAS Code to Perform Predictive Analytics, continued

From the SAS code template that ChatGPT produced, we have pre-constructed code to perform predictive analytics using the MOVIES SAS dataset.

Results:

The LOGISTIC Procedure

Model Information	
Data Set	WORK.MOVIES_PREP
Response Variable	RatingCat
Number of Response Levels	4
Model	generalized logit
Optimization Technique	Newton-Raphson

Number of Observations Read	22
Number of Observations Used	22

Response Profile		
Ordered Value	RatingCat	Total Frequency
1	G	1
2	PG	6
3	PG13	7
4	R	8

Logits modeled use RatingCat='PG' as the reference category.

Class Level Information												
Class	Value	Design Variables										
Category	Action	1	0	0	0	0	0	0	0	0	0	0
	Action Adventure	0	1	0	0	0	0	0	0	0	0	0
	Action Cops & Robber	0	0	1	0	0	0	0	0	0	0	0
	Action Sci-Fi	0	0	0	1	0	0	0	0	0	0	0
	Adventure	0	0	0	0	1	0	0	0	0	0	0
	Comedy	0	0	0	0	0	1	0	0	0	0	0
	Drama	0	0	0	0	0	0	1	0	0	0	0
	Drama Mysteries	0	0	0	0	0	0	0	1	0	0	0
	Drama Romance	0	0	0	0	0	0	0	0	1	0	0
	Drama Suspense	0	0	0	0	0	0	0	0	0	1	0
	Horror	0	0	0	0	0	0	0	0	0	0	1
Studio	Columbia TriStar	1	0	0	0	0	0	0	0	0	0	0
	Filmways Pictures	0	1	0	0	0	0	0	0	0	0	0
	Live Entertainment	0	0	1	0	0	0	0	0	0	0	0
	Lucas Film Ltd	0	0	0	1	0	0	0	0	0	0	0
	MGM / UA	0	0	0	0	1	0	0	0	0	0	0
	Orion	0	0	0	0	0	1	0	0	0	0	0
	Paramount Pictures	0	0	0	0	0	0	1	0	0	0	0
	Universal Pictures	0	0	0	0	0	0	0	1	0	0	0
	Universal Studios	0	0	0	0	0	0	0	0	1	0	0
	Warner Brothers	0	0	0	0	0	0	0	0	0	1	0

Produce SAS Code to Perform Predictive Analytics, continued

Model Convergence Status
Complete separation of data points detected.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	59.991	108.006
SC	63.264	166.922
-2 Log L	53.991	0.006

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	53.9851	51	0.3610
Score	61.4498	51	0.1500
Wald	0.2260	51	1.0000

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Length	3	0.0197	0.9993
Category	18	0.1223	1.0000
Studio	15	0.0381	1.0000
Year	3	0.0002	1.0000

Analysis of Maximum Likelihood Estimates							
Parameter		RatingCat	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		G	1	-347.2	46605.2	0.0001	0.9941
Intercept		PG13	1	-588.6	43361.7	0.0002	0.9892
Intercept		R	1	-176.6	25745.1	0.0000	0.9945
Length		G	1	0.1843	10.2506	0.0003	0.9857
Length		PG13	1	0.2313	8.6438	0.0007	0.9786
Length		R	1	0.3723	2.6547	0.0197	0.8885
Category	Action	G	1	5.6092	1028.0	0.0000	0.9956
Category	Action	PG13	1	5.2424	847.7	0.0000	0.9951
Category	Action	R	1	11.0073	382.8	0.0008	0.9771
Category	Action Adventure	G	1	0.2574	783.3	0.0000	0.9997
Category	Action Adventure	PG13	1	-3.5611	930.1	0.0000	0.9969
Category	Action Adventure	R	1	-0.8164	259.7	0.0000	0.9975
Category	Action Cops & Robber	G	1	8.5323	1007.3	0.0001	0.9932
Category	Action Cops & Robber	PG13	1	-0.0241	943.6	0.0000	1.0000
Category	Action Cops & Robber	R	1	26.5845	322.0	0.0068	0.9342
Category	Action Sci-Fi	G	1	-2.0689	1216.1	0.0000	0.9986
Category	Action Sci-Fi	PG13	1	-10.5871	1164.3	0.0001	0.9927
Category	Action Sci-Fi	R	1	2.3049	425.0	0.0000	0.9957
Category	Adventure	G	1	49.1794	809451	0.0000	1.0000
Category	Adventure	PG13	1	26.1888	1103100	0.0000	1.0000
Category	Adventure	R	1	18.5425	1070807	0.0000	1.0000
Category	Comedy	G	1	12.9594	963.9	0.0002	0.9893
Category	Comedy	PG13	1	14.2180	982.8	0.0002	0.9885
Category	Comedy	R	1	23.5413	281.8	0.0070	0.9334
Category	Drama	G	1	8.2770	846.9	0.0001	0.9922
Category	Drama	PG13	1	14.9548	868.0	0.0003	0.9863
Category	Drama	R	1	5.1275	281.5	0.0003	0.9855

Produce SAS Code to Perform Predictive Analytics, continued

Category	Drama Mysteries	G	1	11.7288	1123.6	0.0001	0.9917
Category	Drama Mysteries	PG13	1	3.7302	1092.9	0.0000	0.9973
Category	Drama Mysteries	R	1	29.8905	377.9	0.0063	0.9370
Category	Drama Romance	G	1	11.1331	900.6	0.0002	0.9901
Category	Drama Romance	PG13	1	19.2145	938.0	0.0004	0.9837
Category	Drama Romance	R	1	1.9448	622.6	0.0000	0.9975
Category	Drama Suspense	G	1	7.5858	1013.6	0.0001	0.9940
Category	Drama Suspense	PG13	1	-2.3927	886.3	0.0000	0.9978
Category	Drama Suspense	R	1	24.3960	339.4	0.0052	0.9427
Category	Horror	G	0	0	.	.	.
Category	Horror	PG13	0	0	.	.	.
Category	Horror	R	0	0	.	.	.
Studio	Columbia TriStar	G	1	5.0562	1007.4	0.0000	0.9960
Studio	Columbia TriStar	PG13	1	-5.7353	881.7	0.0000	0.9947
Studio	Columbia TriStar	R	1	19.8094	338.2	0.0034	0.9533
Studio	Filmways Pictures	G	0	0	.	.	.
Studio	Filmways Pictures	PG13	0	0	.	.	.
Studio	Filmways Pictures	R	0	0	.	.	.
Studio	Live Entertainment	G	1	12.6095	739.0	0.0003	0.9864
Studio	Live Entertainment	PG13	1	12.4904	379.6	0.0011	0.9738
Studio	Live Entertainment	R	1	26.2307	280.2	0.0088	0.9254
Studio	Lucas Film Ltd	G	0	0	.	.	.
Studio	Lucas Film Ltd	PG13	0	0	.	.	.
Studio	Lucas Film Ltd	R	0	0	.	.	.
Studio	MGM / UA	G	1	-1.2036	859.3	0.0000	0.9989
Studio	MGM / UA	PG13	1	-9.9212	1014.4	0.0001	0.9922
Studio	MGM / UA	R	1	5.3580	276.3	0.0004	0.9845
Studio	Orion	G	0	0	.	.	.
Studio	Orion	PG13	0	0	.	.	.
Studio	Orion	R	0	0	.	.	.
Studio	Paramount Pictures	G	1	-5.5912	570.4	0.0001	0.9922
Studio	Paramount Pictures	PG13	1	-14.4199	240.1	0.0036	0.9521
Studio	Paramount Pictures	R	1	0.9671	181.3	0.0000	0.9957
Studio	Universal Pictures	G	0	0	.	.	.
Studio	Universal Pictures	PG13	0	0	.	.	.
Studio	Universal Pictures	R	0	0	.	.	.
Studio	Universal Studios	G	1	-3.0117	817.6	0.0000	0.9971
Studio	Universal Studios	PG13	1	-6.4573	450.7	0.0002	0.9886
Studio	Universal Studios	R	1	-14.9602	816.3	0.0003	0.9854
Studio	Warner Brothers	G	0	0	.	.	.
Studio	Warner Brothers	PG13	0	0	.	.	.
Studio	Warner Brothers	R	0	0	.	.	.
Year		G	1	0.1588	23.4111	0.0000	0.9946
Year		PG13	1	0.2832	21.5155	0.0002	0.9895
Year		R	1	0.0595	13.0168	0.0000	0.9964

Produce SAS Code to Perform Predictive Analytics, continued

Odds Ratio Estimates				
Effect	RatingCat	Point Estimate	95% Wald Confidence Limits	
Length	G	1.202	<0.001	>999.999
Length	PG13	1.260	<0.001	>999.999
Length	R	1.451	0.008	263.874
Category Action vs Horror	G	272.927	<0.001	>999.999
Category Action vs Horror	PG13	189.121	<0.001	>999.999
Category Action vs Horror	R	>999.999	<0.001	>999.999
Category Action Adventure vs Horror	G	1.294	<0.001	>999.999
Category Action Adventure vs Horror	PG13	0.028	<0.001	>999.999
Category Action Adventure vs Horror	R	0.442	<0.001	>999.999
Category Action Cops & Robber vs Horror	G	>999.999	<0.001	>999.999
Category Action Cops & Robber vs Horror	PG13	0.976	<0.001	>999.999
Category Action Cops & Robber vs Horror	R	>999.999	<0.001	>999.999
Category Action Sci-Fi vs Horror	G	0.126	<0.001	>999.999
Category Action Sci-Fi vs Horror	PG13	<0.001	<0.001	>999.999
Category Action Sci-Fi vs Horror	R	10.023	<0.001	>999.999
Category Adventure vs Horror	G	>999.999	<0.001	>999.999
Category Adventure vs Horror	PG13	>999.999	<0.001	>999.999
Category Adventure vs Horror	R	>999.999	<0.001	>999.999
Category Comedy vs Horror	G	>999.999	<0.001	>999.999
Category Comedy vs Horror	PG13	>999.999	<0.001	>999.999
Category Comedy vs Horror	R	>999.999	<0.001	>999.999
Category Drama vs Horror	G	>999.999	<0.001	>999.999
Category Drama vs Horror	PG13	>999.999	<0.001	>999.999
Category Drama vs Horror	R	168.597	<0.001	>999.999
Category Drama Mysteries vs Horror	G	>999.999	<0.001	>999.999
Category Drama Mysteries vs Horror	PG13	41.686	<0.001	>999.999
Category Drama Mysteries vs Horror	R	>999.999	<0.001	>999.999
Category Drama Romance vs Horror	G	>999.999	<0.001	>999.999
Category Drama Romance vs Horror	PG13	>999.999	<0.001	>999.999
Category Drama Romance vs Horror	R	6.992	<0.001	>999.999
Category Drama Suspense vs Horror	G	>999.999	<0.001	>999.999
Category Drama Suspense vs Horror	PG13	0.091	<0.001	>999.999
Category Drama Suspense vs Horror	R	>999.999	<0.001	>999.999
Studio Columbia TriStar vs Warner Brothers	G	156.988	<0.001	>999.999
Studio Columbia TriStar vs Warner Brothers	PG13	0.003	<0.001	>999.999
Studio Columbia TriStar vs Warner Brothers	R	>999.999	<0.001	>999.999
Studio Live Entertainment vs Warner Brothers	G	>999.999	<0.001	>999.999
Studio Live Entertainment vs Warner Brothers	PG13	>999.999	<0.001	>999.999
Studio Live Entertainment vs Warner Brothers	R	>999.999	<0.001	>999.999
Studio MGM / UA vs Warner Brothers	G	0.300	<0.001	>999.999
Studio MGM / UA vs Warner Brothers	PG13	<0.001	<0.001	>999.999
Studio MGM / UA vs Warner Brothers	R	212.298	<0.001	>999.999
Studio Paramount Pictures vs Warner Brothers	G	0.004	<0.001	>999.999

Produce SAS Code to Perform Predictive Analytics, continued

Studio Paramount Pictures vs Warner Brothers	PG13	<0.001	<0.001	>999.999
Studio Paramount Pictures vs Warner Brothers	R	2.630	<0.001	>999.999
Studio Universal Studios vs Warner Brothers	G	0.049	<0.001	>999.999
Studio Universal Studios vs Warner Brothers	PG13	0.002	<0.001	>999.999
Studio Universal Studios vs Warner Brothers	R	<0.001	<0.001	>999.999
Year	G	1.172	<0.001	>999.999
Year	PG13	1.327	<0.001	>999.999
Year	R	1.061	<0.001	>999.999

The HPSPLIT Procedure

Performance Information

Execution Mode	Single-Machine
Number of Threads	2

Data Access Information

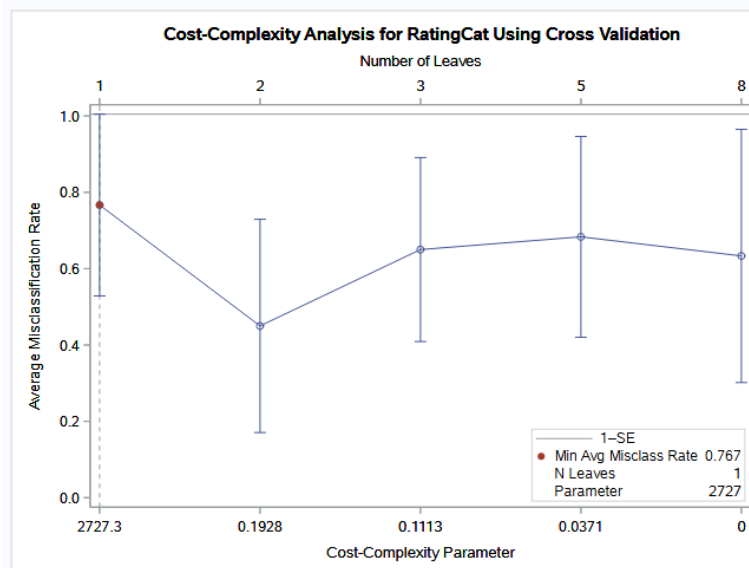
Data	Engine	Role	Path
WORK.MOVIES_PREP	V9	Input	On Client

Model Information

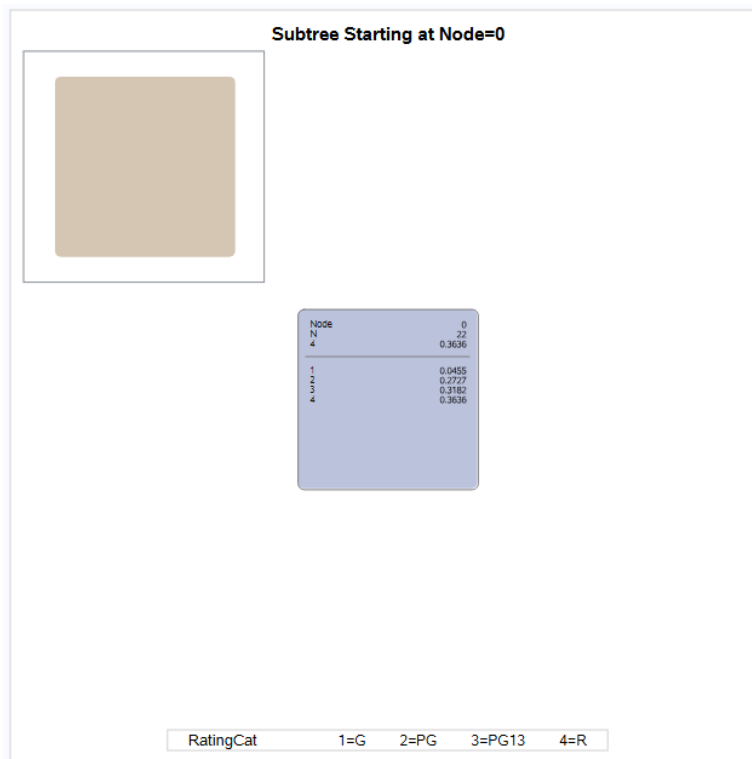
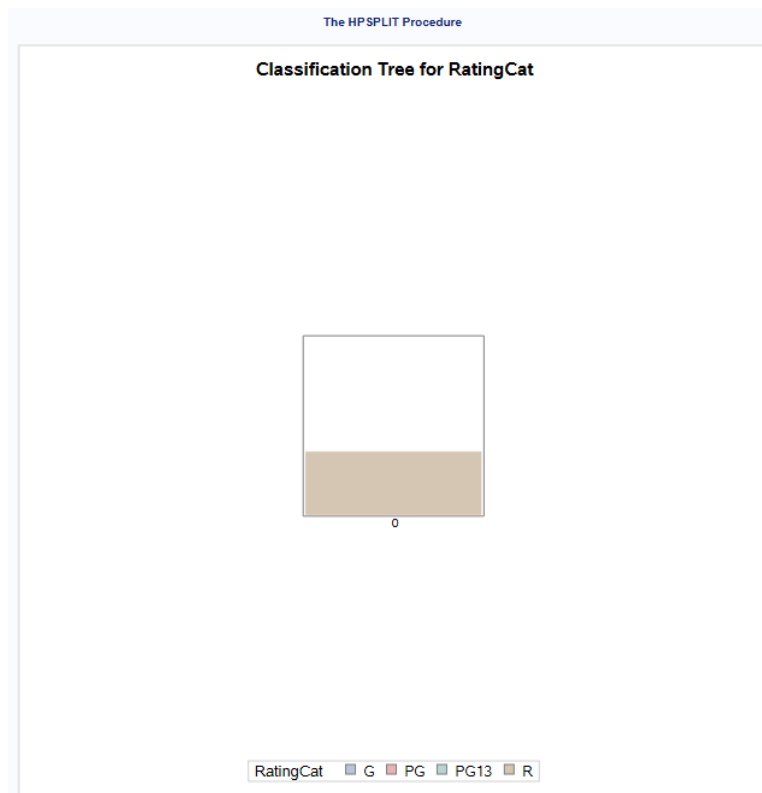
Split Criterion Used	Entropy
Pruning Method	Cost-Complexity
Subtree Evaluation Criterion	Cost-Complexity
Number of Branches	2
Maximum Tree Depth Requested	10
Maximum Tree Depth Achieved	4
Tree Depth	0
Number of Leaves Before Pruning	8
Number of Leaves After Pruning	1

Number of Observations Read	22
Number of Observations Used	22

The HPSPLIT Procedure



Produce SAS Code to Perform Predictive Analytics, continued



Produce SAS Code to Perform Predictive Analytics, continued

The HPSPPLIT Procedure

Model-Based Confusion Matrix					
Actual	Predicted				Error Rate
	G	PG	PG13	R	
G	0	0	0	1	1.0000
PG	0	0	0	6	1.0000
PG13	0	0	0	7	1.0000
R	0	0	0	8	0.0000

Model-Based Fit Statistics for Selected Tree

N Leaves	ASE	Mis-class	Entropy	Gini	RSS
1	0.1725	0.6364	1.7703	0.6901	15.1818

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling
------------------	------------------------

Input Data Set	MOVIES_PREP
Random Number Seed	12345
Sampling Rate	0.7
Sample Size	16
Selection Probability	0.727273
Sampling Weight	0
Output Data Set	MOVIES_SPLIT

The HPSPPLIT Procedure

Performance Information

Execution Mode	Single-Machine
Number of Threads	2

Data Access Information

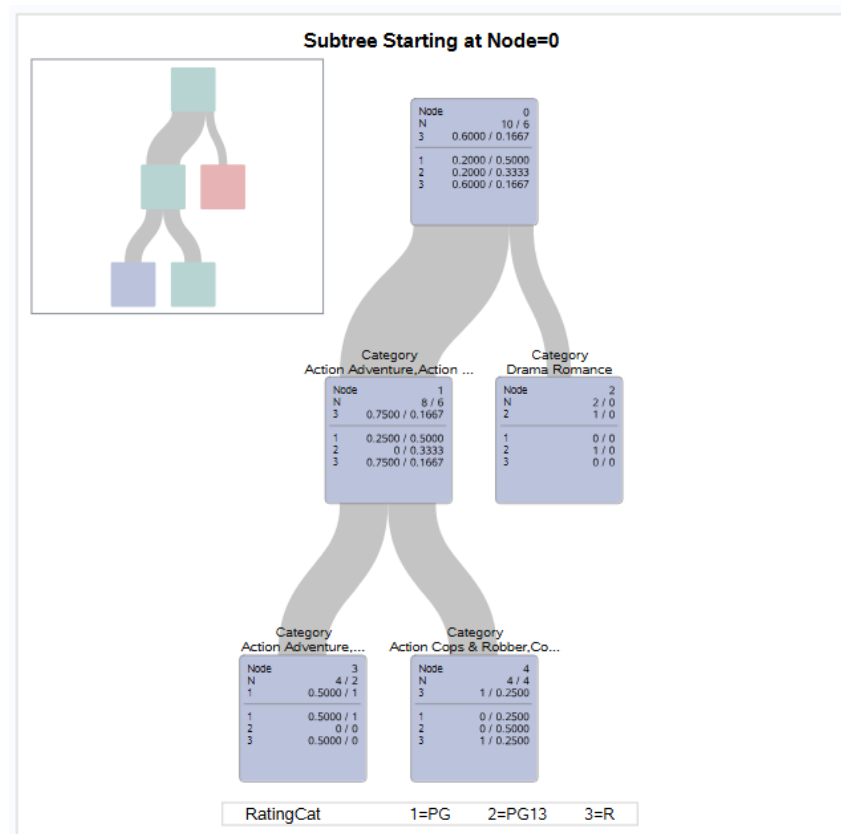
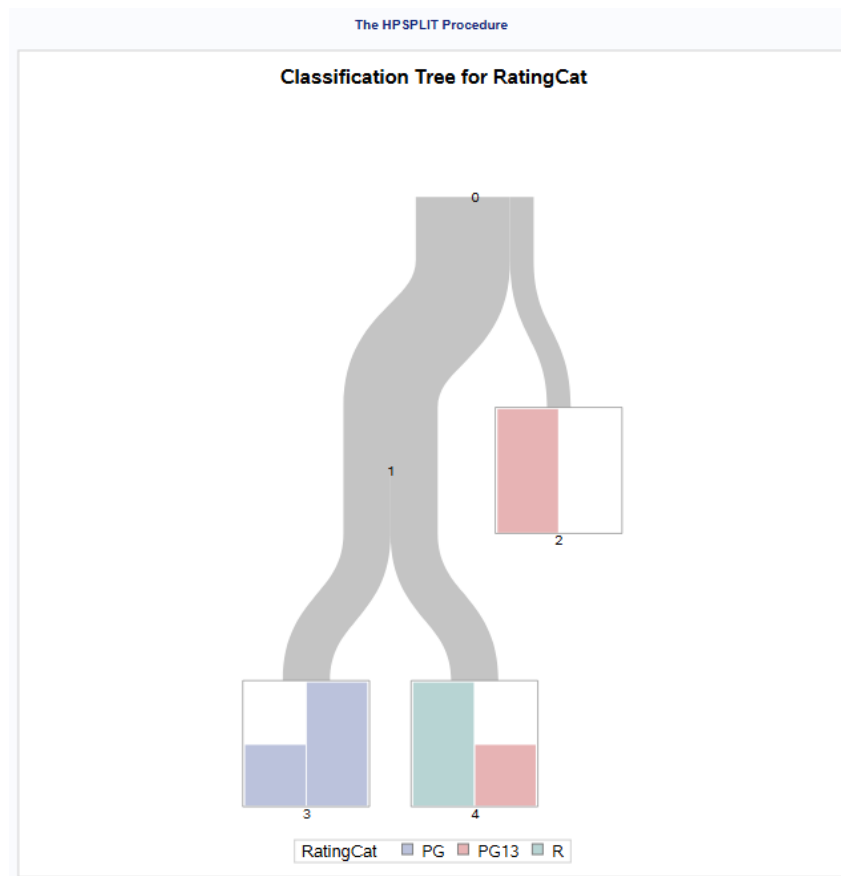
Data	Engine	Role	Path
WORK.TRAIN	V9	Input	On Client
WORK.TREE_OUT	V9	Output	On Client

Model Information

Split Criterion Used	Entropy
Pruning Method	Cost-Complexity
Subtree Evaluation Criterion	Cost-Complexity
Number of Branches	2
Maximum Tree Depth Requested	10
Maximum Tree Depth Achieved	3
Tree Depth	2
Number of Leaves Before Pruning	4
Number of Leaves After Pruning	3

Number of Observations Read	16
Number of Observations Used	16
Number of Training Observations Used	10
Number of Validation Observations Used	6

Produce SAS Code to Perform Predictive Analytics, continued



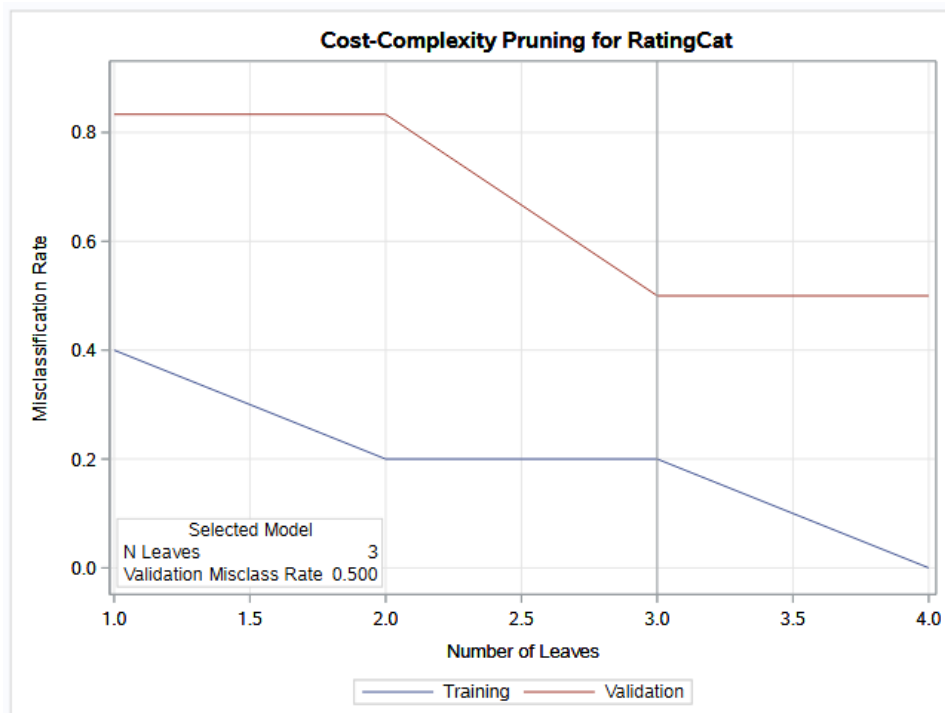
Produce SAS Code to Perform Predictive Analytics, continued

The HPSPLIT Procedure

Confusion Matrices					
	Actual	Predicted			Error Rate
		PG	PG13	R	
Training	PG	2	0	0	0.0000
	PG13	0	2	0	0.0000
	R	2	0	4	0.3333
Validation	PG	2	0	1	0.3333
	PG13	0	0	2	1.0000
	R	0	0	1	0.0000

Fit Statistics for Selected Tree						
	N Leaves	ASE	Mis-class	Entropy	Gini	RSS
Training	3	0.0667	0.2000	0.4000	0.2000	2.0000
Validation	3	0.3889	0.5000	1.0000	0.4167	7.0000

Variable Importance						
Variable	Training		Validation		Relative Ratio	Count
	Relative	Importance	Relative	Importance		
Category	1.0000	1.8974		0		2



Produce SAS Code to Perform Predictive Analytics, continued

Confusion Matrix: Actual vs Predicted Ratings

The FREQ Procedure

Frequency	Table of RatingCat by P_RatingCatPG		
	P_RatingCatPG(Predicted: RatingCat=PG)		Total
	RatingCat	0	
	G	1	1
	PG	1	1
	PG13	3	3
	R	1	1
	Total	6	6

Frequency	Table of RatingCat by P_RatingCatPG13		
	P_RatingCatPG13(Predicted: RatingCat=PG13)		Total
	RatingCat	0	
	G	1	1
	PG	1	1
	PG13	3	3
	R	1	1
	Total	6	6

Frequency	Table of RatingCat by P_RatingCatR		
	P_RatingCatR(Predicted: RatingCat=R)		Total
	RatingCat	1	
	G	1	1
	PG	1	1
	PG13	3	3
	R	1	1
	Total	6	6

ROC Curve: Predicting G Rating vs Others

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TRAIN
Response Variable	RatingCat
Number of Response Levels	3
Model	cumulative logit
Optimization Technique	Fisher's scoring

Number of Observations Read	16
Number of Observations Used	16

Response Profile		
Ordered Value	RatingCat	Total Frequency
1	PG	5
2	PG13	4
3	R	7

Produce SAS Code to Perform Predictive Analytics, continued

Probabilities modeled are cumulated over the lower Ordered Values.

Class Level Information									
Class	Value	Design Variables							
Category	Action Adventure	1	0	0	0	0	0	0	0
	Action Cops & Robber	0	1	0	0	0	0	0	0
	Comedy	0	0	1	0	0	0	0	0
	Drama	0	0	0	1	0	0	0	0
	Drama Mysteries	0	0	0	0	1	0	0	0
	Drama Romance	0	0	0	0	0	1	0	0
	Drama Suspense	0	0	0	0	0	0	1	0
	Horror	0	0	0	0	0	0	0	1
Studio	Columbia TriStar	1	0	0	0	0	0	0	
	Filmways Pictures	0	1	0	0	0	0	0	
	MGM / UA	0	0	1	0	0	0	0	
	Orion	0	0	0	1	0	0	0	
	Paramount Pictures	0	0	0	0	1	0	0	
	Universal Studios	0	0	0	0	0	1	0	
	Warner Brothers	0	0	0	0	0	0	1	

Model Convergence Status

Quasi-complete separation of data points detected.

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
6.5404	13	0.9243

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	38.295	38.035
SC	39.841	49.624
-2 Log L	34.295	8.035

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	26.2605	13	0.0157
Score	13.8715	13	0.3830
Wald	1.2683	13	1.0000

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Length	1	0.1458	0.7026
Category	5	0.2148	0.9989
Studio	4	0.0536	0.9996
Year	1	0.3188	0.5723

Produce SAS Code to Perform Predictive Analytics, continued

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	PG	1	863.2	1507.7	0.3278	0.5670
Intercept	PG13	1	868.4	1510.3	0.3306	0.5653
Length		1	-0.0272	0.0713	0.1458	0.7026
Category	Action Adventure	1	-2.4236	114.2	0.0005	0.9831
Category	Action Cops & Robber	1	-25.8370	138.4	0.0348	0.8519
Category	Comedy	1	-15.9112	127.1	0.0157	0.9004
Category	Drama	1	-17.3909	117.9	0.0218	0.8827
Category	Drama Mysteries	1	-29.7796	153.0	0.0379	0.8457
Category	Drama Romance	1	-1.6612	114.2	0.0002	0.9884
Category	Drama Suspense	1	-24.7331	152.5	0.0263	0.8712
Category	Horror	0	0	.	.	.
Studio	Columbia TriStar	1	-23.5533	152.5	0.0239	0.8772
Studio	Filmways Pictures	0	0	.	.	.
Studio	MGM / UA	1	-5.7168	98.4426	0.0034	0.9537
Studio	Orion	0	0	.	.	.
Studio	Paramount Pictures	1	-9.3358	55.5332	0.0283	0.8665
Studio	Universal Studios	1	-3.6888	209.6	0.0003	0.9860
Studio	Warner Brothers	0	0	.	.	.
Year		1	-0.4266	0.7555	0.3188	0.5723

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Length	0.973	0.846	1.119
Category Action Adventure vs Horror	0.089	<0.001	>999.999
Category Action Cops & Robber vs Horror	<0.001	<0.001	>999.999
Category Comedy vs Horror	<0.001	<0.001	>999.999
Category Drama vs Horror	<0.001	<0.001	>999.999
Category Drama Mysteries vs Horror	<0.001	<0.001	>999.999
Category Drama Romance vs Horror	0.190	<0.001	>999.999
Category Drama Suspense vs Horror	<0.001	<0.001	>999.999
Studio Columbia TriStar vs Warner Brothers	<0.001	<0.001	>999.999
Studio MGM / UA vs Warner Brothers	0.003	<0.001	>999.999
Studio Paramount Pictures vs Warner Brothers	<0.001	<0.001	>999.999
Studio Universal Studios vs Warner Brothers	0.025	<0.001	>999.999
Year	0.653	0.148	2.870

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	97.6	Somers' D	0.952
Percent Discordant	2.4	Gamma	0.952
Percent Tied	0.0	Tau-a	0.658
Pairs	83	c	0.976

Produce SAS Code to Perform Predictive Analytics, continued

ROC Curve: Predicting PG-13 Rating vs Others

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TRAIN
Response Variable	RatingCat
Number of Response Levels	3
Model	cumulative logit
Optimization Technique	Fisher's scoring

Number of Observations Read	16
Number of Observations Used	16

Response Profile		
Ordered Value	RatingCat	Total Frequency
1	PG	5
2	PG13	4
3	R	7

Probabilities modeled are cumulated over the lower Ordered Values.

Class Level Information									
Class	Value	Design Variables							
Category	Action Adventure	1	0	0	0	0	0	0	0
	Action Cops & Robber	0	1	0	0	0	0	0	0
	Comedy	0	0	1	0	0	0	0	0
	Drama	0	0	0	1	0	0	0	0
	Drama Mysteries	0	0	0	0	1	0	0	0
	Drama Romance	0	0	0	0	0	1	0	0
	Drama Suspense	0	0	0	0	0	0	1	0
Studio	Horror	0	0	0	0	0	0	0	1
	Columbia TriStar	1	0	0	0	0	0	0	0
	Filmways Pictures	0	1	0	0	0	0	0	0
	MGM / UA	0	0	1	0	0	0	0	0
	Orion	0	0	0	1	0	0	0	0
	Paramount Pictures	0	0	0	0	1	0	0	0
	Universal Studios	0	0	0	0	0	1	0	0
	Warner Brothers	0	0	0	0	0	0	1	0

Produce SAS Code to Perform Predictive Analytics, continued

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	PG	1	863.2	1507.7	0.3278	0.5670
Intercept	PG13	1	868.4	1510.3	0.3306	0.5653
Length		1	-0.0272	0.0713	0.1458	0.7026
Category	Action Adventure	1	-2.4236	114.2	0.0005	0.9831
Category	Action Cops & Robber	1	-25.8370	138.4	0.0348	0.8519
Category	Comedy	1	-15.9112	127.1	0.0157	0.9004
Category	Drama	1	-17.3909	117.9	0.0218	0.8827
Category	Drama Mysteries	1	-29.7796	153.0	0.0379	0.8457
Category	Drama Romance	1	-1.6612	114.2	0.0002	0.9884
Category	Drama Suspense	1	-24.7331	152.5	0.0263	0.8712
Category	Horror	0	0	.	.	.
Studio	Columbia TriStar	1	-23.5533	152.5	0.0239	0.8772
Studio	Filmways Pictures	0	0	.	.	.
Studio	MGM / UA	1	-5.7168	98.4426	0.0034	0.9537
Studio	Orion	0	0	.	.	.
Studio	Paramount Pictures	1	-9.3358	55.5332	0.0283	0.8665
Studio	Universal Studios	1	-3.6888	209.6	0.0003	0.9860
Studio	Warner Brothers	0	0	.	.	.
Year		1	-0.4266	0.7555	0.3188	0.5723

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Length	0.973	0.846	1.119
Category Action Adventure vs Horror	0.089	<0.001	>999.999
Category Action Cops & Robber vs Horror	<0.001	<0.001	>999.999
Category Comedy vs Horror	<0.001	<0.001	>999.999
Category Drama vs Horror	<0.001	<0.001	>999.999
Category Drama Mysteries vs Horror	<0.001	<0.001	>999.999
Category Drama Romance vs Horror	0.190	<0.001	>999.999
Category Drama Suspense vs Horror	<0.001	<0.001	>999.999
Studio Columbia TriStar vs Warner Brothers	<0.001	<0.001	>999.999
Studio MGM / UA vs Warner Brothers	0.003	<0.001	>999.999
Studio Paramount Pictures vs Warner Brothers	<0.001	<0.001	>999.999
Studio Universal Studios vs Warner Brothers	0.025	<0.001	>999.999
Year	0.653	0.148	2.870

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	97.6	Somers' D	0.952
Percent Discordant	2.4	Gamma	0.952
Percent Tied	0.0	Tau-a	0.658
Pairs	83	c	0.976

Produce SAS Code to Perform Predictive Analytics, continued

ROC Curve: Predicting R Rating vs Others

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TRAIN
Response Variable	RatingCat
Number of Response Levels	3
Model	cumulative logit
Optimization Technique	Fisher's scoring

Number of Observations Read	16
Number of Observations Used	16

Response Profile		
Ordered Value	RatingCat	Total Frequency
1	PG	5
2	PG13	4
3	R	7

Probabilities modeled are cumulated over the lower Ordered Values.

Class Level Information										
Class	Value	Design Variables								
Category	Action Adventure	1	0	0	0	0	0	0	0	0
	Action Cops & Robber	0	1	0	0	0	0	0	0	0
	Comedy	0	0	1	0	0	0	0	0	0
	Drama	0	0	0	1	0	0	0	0	0
	Drama Mysteries	0	0	0	0	1	0	0	0	0
	Drama Romance	0	0	0	0	0	1	0	0	0
	Drama Suspense	0	0	0	0	0	0	1	0	0
	Horror	0	0	0	0	0	0	0	0	1
Studio	Columbia TriStar	1	0	0	0	0	0	0	0	
	Filmways Pictures	0	1	0	0	0	0	0	0	
	MGM / UA	0	0	1	0	0	0	0	0	
	Orion	0	0	0	1	0	0	0	0	
	Paramount Pictures	0	0	0	0	1	0	0	0	
	Universal Studios	0	0	0	0	0	1	0	0	
	Warner Brothers	0	0	0	0	0	0	1	0	

Produce SAS Code to Perform Predictive Analytics, continued

Score Test for the Proportional Odds Assumption		
Chi-Square	DF	Pr > ChiSq
6.5404	13	0.9243

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	38.295	38.035
SC	39.841	49.624
-2 Log L	34.295	8.035

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	26.2605	13	0.0157
Score	13.8715	13	0.3830
Wald	1.2683	13	1.0000

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Length	1	0.1458	0.7026
Category	5	0.2148	0.9989
Studio	4	0.0536	0.9996
Year	1	0.3188	0.5723

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	PG	1	863.2	1507.7	0.3278	0.5670
Intercept	PG13	1	868.4	1510.3	0.3306	0.5653
Length		1	-0.0272	0.0713	0.1458	0.7026
Category	Action Adventure	1	-2.4236	114.2	0.0005	0.9831
Category	Action Cops & Robber	1	-25.8370	138.4	0.0348	0.8519
Category	Comedy	1	-15.9112	127.1	0.0157	0.9004
Category	Drama	1	-17.3909	117.9	0.0218	0.8827
Category	Drama Mysteries	1	-29.7796	153.0	0.0379	0.8457
Category	Drama Romance	1	-1.6612	114.2	0.0002	0.9884
Category	Drama Suspense	1	-24.7331	152.5	0.0263	0.8712
Category	Horror	0	0	.	.	.
Studio	Columbia TriStar	1	-23.5533	152.5	0.0239	0.8772
Studio	Filmways Pictures	0	0	.	.	.
Studio	MGM / UA	1	-5.7168	98.4426	0.0034	0.9537
Studio	Orion	0	0	.	.	.
Studio	Paramount Pictures	1	-9.3358	55.5332	0.0283	0.8665
Studio	Universal Studios	1	-3.6888	209.6	0.0003	0.9880
Studio	Warner Brothers	0	0	.	.	.
Year		1	-0.4266	0.7555	0.3188	0.5723

Produce SAS Code to Perform Predictive Analytics, continued

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Length	0.973	0.846	1.119
Category Action Adventure vs Horror	0.089	<0.001	>999.999
Category Action Cops & Robber vs Horror	<0.001	<0.001	>999.999
Category Comedy vs Horror	<0.001	<0.001	>999.999
Category Drama vs Horror	<0.001	<0.001	>999.999
Category Drama Mysteries vs Horror	<0.001	<0.001	>999.999
Category Drama Romance vs Horror	0.190	<0.001	>999.999
Category Drama Suspense vs Horror	<0.001	<0.001	>999.999
Studio Columbia TriStar vs Warner Brothers	<0.001	<0.001	>999.999
Studio MGM / UA vs Warner Brothers	0.003	<0.001	>999.999
Studio Paramount Pictures vs Warner Brothers	<0.001	<0.001	>999.999
Studio Universal Studios vs Warner Brothers	0.025	<0.001	>999.999
Year	0.653	0.148	2.870

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	97.6	Somers' D	0.952
Percent Discordant	2.4	Gamma	0.952
Percent Tied	0.0	Tau-a	0.658
Pairs	83	c	0.976

Variable Importance from Decision Tree

The HPSPLIT Procedure

Performance Information

Execution Mode	Single-Machine
Number of Threads	2

Data Access Information

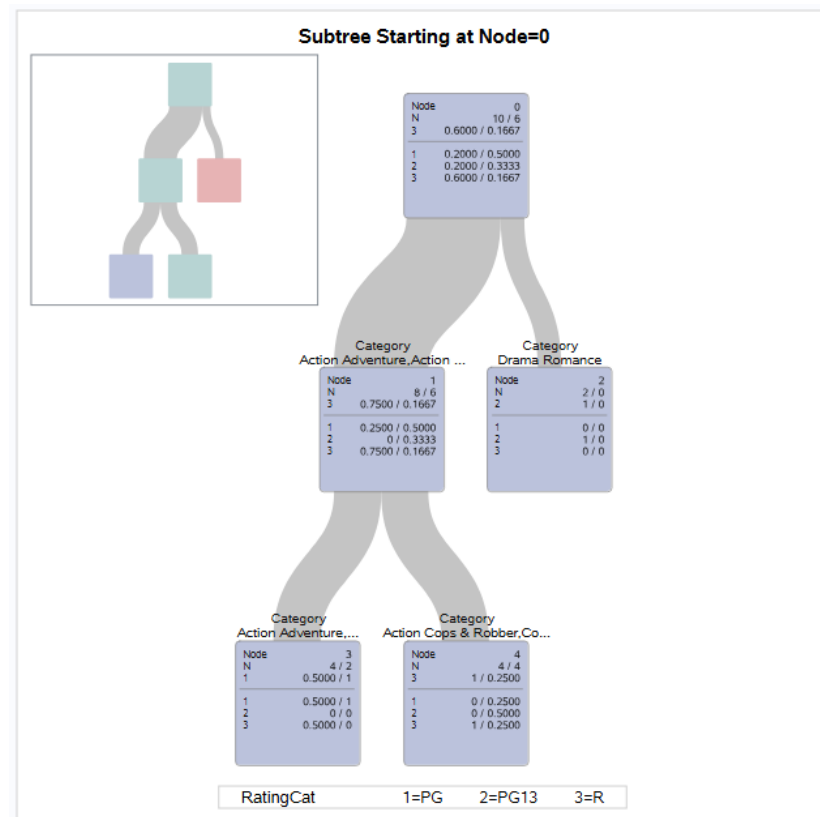
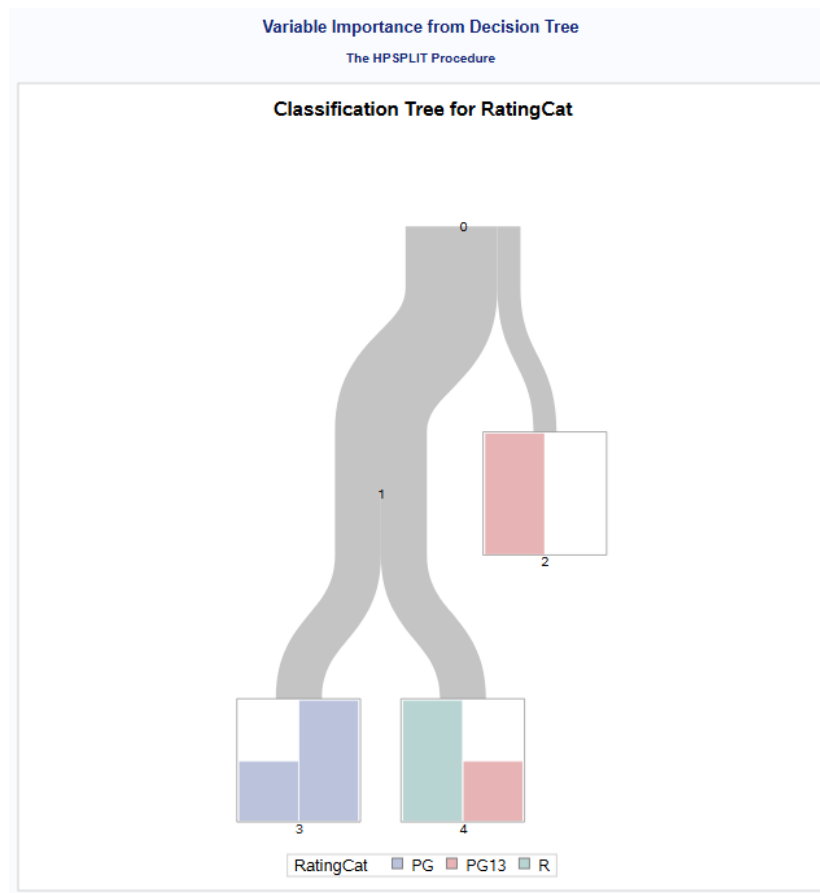
Data	Engine	Role	Path
WORK.TRAIN	V9	Input	On Client
WORK.TREE_OUT	V9	Output	On Client

Model Information

Split Criterion Used	Entropy
Pruning Method	Cost-Complexity
Subtree Evaluation Criterion	Cost-Complexity
Number of Branches	2
Maximum Tree Depth Requested	10
Maximum Tree Depth Achieved	3
Tree Depth	2
Number of Leaves Before Pruning	4
Number of Leaves After Pruning	3

Number of Observations Read	16
Number of Observations Used	16
Number of Training Observations Used	10
Number of Validation Observations Used	6

Produce SAS Code to Perform Predictive Analytics, continued



Produce SAS Code to Perform Predictive Analytics, continued

Variable Importance from Decision Tree

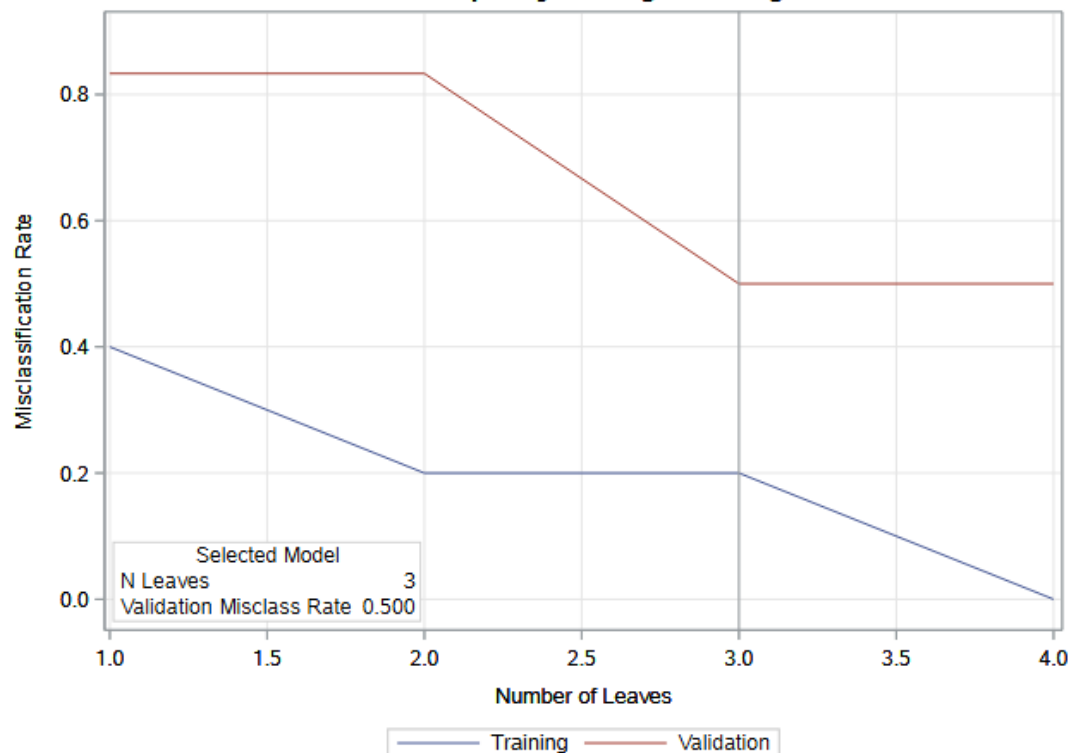
The HPSPLOT Procedure

Confusion Matrices					
	Actual	Predicted			Error Rate
		PG	PG13	R	
Training	PG	2	0	0	0.0000
	PG13	0	2	0	0.0000
	R	2	0	4	0.3333
Validation	PG	2	0	1	0.3333
	PG13	0	0	2	1.0000
	R	0	0	1	0.0000

Fit Statistics for Selected Tree						
	N Leaves	ASE	Mis-class	Entropy	Gini	RSS
Training	3	0.0667	0.2000	0.4000	0.2000	2.0000
Validation	3	0.3889	0.5000	1.0000	0.4167	7.0000

Variable Importance						
Variable	Training		Validation		Relative Ratio	Count
	Relative	Importance	Relative	Importance		
Category	1.0000	1.8974		0		2

Cost-Complexity Pruning for RatingCat



CONCLUSION

Incorporating AI into our analytics workflows isn't about replacing human insight—it's about enhancing it. By leveraging the structured power of SAS alongside the flexible intelligence of ChatGPT, we unlock new levels of efficiency, clarity, and creativity. The "5 C's" framework offers a roadmap for doing more with less: automating where we can, customizing where we need to, and collaborating in ways that elevate team performance. As the analytics landscape continues to evolve, the partnership between human expertise, statistical rigor, and conversational AI will be a defining factor in driving meaningful, impactful results. With SAS and ChatGPT working together, we are not just getting the job done—we're redefining how it's done.

REFERENCES

- Caswell, Amanda. (2025, April 13). I use ChatGPT every day — here's 10 prompts I can't live without. Tom's Guide. Retrieved: April 28, 2025 from <https://www.tomsguide.com/ai/i-use-chatgpt-every-day-heres-10-prompts-i-cant-live-without>.
- Horton, Cody L. (2025, April 23). Maximizing Productivity with ChatGPT and Generative AI Tools. Cody L. Horton. Retrieved: May 1, 2025 from <https://codylhorton.com/maximizing-productivity-with-chatgpt-and-generative-ai-tools/>.
- Hreha, Jason. (2024, March 21). 19 Ways to Use ChatGPT to Boost Your Productivity. Persona. Retrieved: April 28, 2025 from <https://www.personalent.com/productivity/ways-to-use-chatgpt-to-boost-your-productivity/>.
- Lafler, Kirk Paul. (2019). *PROC SQL: Beyond the Basics Using SAS, Third Edition*, SAS Institute Inc., Cary, NC, USA.
- Lulka, Jess. (2025, April 18). 10 ChatGPT Productivity Hacks to Boost Efficiency and Save Time. DigitalOcean, Retrieved: April 28, 2025 from <https://www.digitalocean.com/resources/articles/chatgpt-productivity-hacks>.
- Rosenthal, Joel. (2025, April 14). How to Use ChatGPT for Answers, Inspiration & Productivity. Charter Capital. Retrieved: May 1, 2025 from <https://www.charcap.com/chatgpt-business/>.

ACKNOWLEDGMENTS

The authors thank the PharmaSUG 2025 Conference Committee, particularly the PharmaSUG 2025 Artificial Intelligence and Machine Learning (AI) Section Chairs, Bradford Danner and Venky Chakravarthy, for accepting our abstract and paper; the PharmaSUG 2025 Academic Chair, Ajay Gupta, the PharmaSUG 2025 Operations Chair, Gary Moore, for organizing and supporting a great "in-person" conference event; SAS Institute Inc. for providing SAS users with wonderful software; and SAS users everywhere for being the nicest people anywhere!

RECOMMENDED READING

- *Base SAS® Procedures Guide*
- *SAS® For Dummies®*

CONTACT INFORMATION

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

Charu Shankar
SAS Institute Toronto
EMAIL: Charu.shankar@sas.com
BLOG: <https://blogs.sas.com/content/author/charushankar/>
TWITTER: CharuYogaCan
LINKEDIN: <https://www.linkedin.com/in/charushankar/>

~ ~ ~ ~ ~

Kirk Paul Lafler
Data Scientist, Consultant, Developer, Programmer, Educator, and Author
Specializing in SAS® / Python / SQL / DBMS / Excel / R / AWS / Cloud-based Technologies
EMAIL: KirkLafler@cs.com
LINKEDIN: <https://www.linkedin.com/in/KirkPaulLafler/>

Any brand and product names are trademarks of their respective companies.