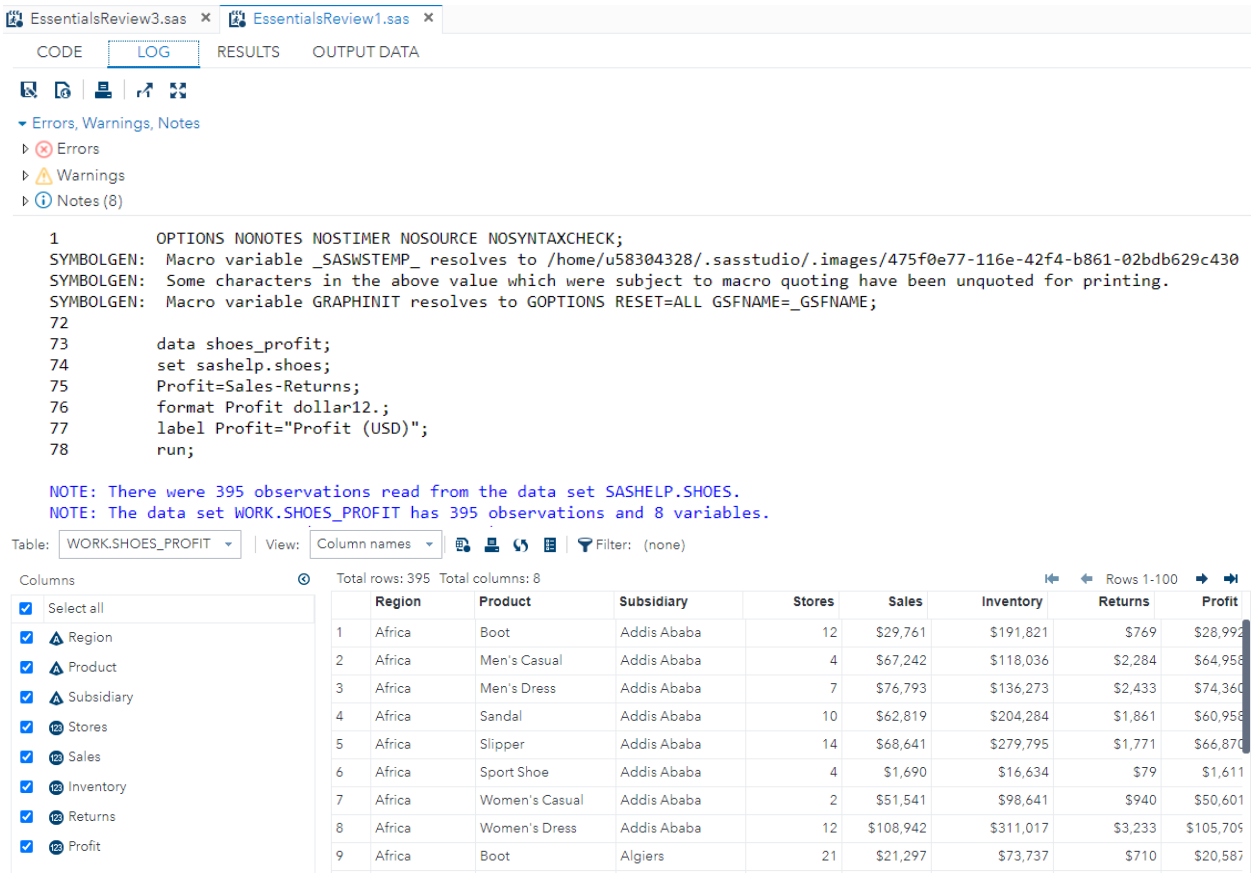
Preparing for the SAS Programming Certification

Week 1: Essential Reviews, Accessing Data Review, Exploring Data Review

```
/*********************************************/
/* This code defines macro variables and the   */
/* library for this course. You must run       */
/* this code each time you start SAS OnDemand  */
/* for Academics to access your practice data. */
/*********************************************/

%let path=~/ECRB94/data;
%let outpath=~/ECRB94/output;
libname cr "&path";

*Essentials Review 1;
data shoes_profit;
        set sashelp.shoes;
        Profit=Sales-Returns;
        format Profit dollar12.;
        label Profit="Profit (USD)";
run;

proc means data=shoes_profit noprint;
        var Profit;
        class Region Subsidiary;
        output out=shoes_summary sum=Profit;
        ways 2;
run;
```

```
title "Shoes Profit by Region, Subsidiary";
proc print data=shoes_summary noobs;
        by Region;
        var Subsidiary Profit;
run;
title;
```



EssentialsReview3.sas × EssentialsReview1.sas ×

CODE LOG RESULTS OUTPUT DATA

Errors, Warnings, Notes
- Errors
- Warnings
- Notes (8)

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
SYMBOLGEN:  Macro variable _SASWSTEMP_ resolves to /home/u58304328/.sasstudio/.images/475f0e77-116e-42f4-b861-02bdb629c430
SYMBOLGEN:  Some characters in the above value which were subject to macro quoting have been unquoted for printing.
SYMBOLGEN:  Macro variable GRAPHINIT resolves to GOPTIONS RESET=ALL GSFNAME=_GSFNAME;
72
73         data shoes_profit;
74         set sashelp.shoes;
75         Profit=Sales-Returns;
76         format Profit dollar12.;
77         label Profit="Profit (USD)";
78         run;

NOTE: There were 395 observations read from the data set SASHELP.SHOES.
NOTE: The data set WORK.SHOES_PROFIT has 395 observations and 8 variables.
```

Table: WORK.SHOES_PROFIT | View: Column names | Filter: (none)

Columns: Select all, Region, Product, Subsidiary, Stores, Sales, Inventory, Returns, Profit

Total rows: 395 Total columns: 8 — Rows 1-100

	Region	Product	Subsidiary	Stores	Sales	Inventory	Returns	Profit
1	Africa	Boot	Addis Ababa	12	$29,761	$191,821	$769	$28,992
2	Africa	Men's Casual	Addis Ababa	4	$67,242	$118,036	$2,284	$64,958
3	Africa	Men's Dress	Addis Ababa	7	$76,793	$136,273	$2,433	$74,360
4	Africa	Sandal	Addis Ababa	10	$62,819	$204,284	$1,861	$60,958
5	Africa	Slipper	Addis Ababa	14	$68,641	$279,795	$1,771	$66,870
6	Africa	Sport Shoe	Addis Ababa	4	$1,690	$16,634	$79	$1,611
7	Africa	Women's Casual	Addis Ababa	2	$51,541	$98,641	$940	$50,601
8	Africa	Women's Dress	Addis Ababa	12	$108,942	$311,017	$3,233	$105,709
9	Africa	Boot	Algiers	21	$21,297	$73,737	$710	$20,587

Table: WORK.SHOES_SUMMARY | View: Column names | Filter: (none)

Columns

- Select all
- Region
- Subsidiary
- _TYPE_
- _FREQ_
- Profit

Total rows: 53 Total columns: 5

	Region	Subsidiary	_TYPE_	_FREQ_	Profit
1	Africa	Addis Ababa	3	8	$454,059
2	Africa	Algiers	3	7	$382,837
3	Africa	Cairo	3	8	$715,721
4	Africa	Johannesburg	3	5	$109,046
5	Africa	Khartoum	3	8	$179,541
6	Africa	Kinshasa	3	7	$191,234
7	Africa	Luanda	3	7	$133,787

Shoes Profit by Region, Subsidiary

Region=Africa

Subsidiary	Profit
Addis Ababa	$454,059
Algiers	$382,837
Cairo	$715,721
Johannesburg	$109,046
Khartoum	$179,541
Kinshasa	$191,234
Luanda	$133,787
Nairobi	$102,276

Region=Asia

Subsidiary	Profit
Bangkok	$16,103
Seoul	$432,100
Tokyo	$1,133

Region=Canada

Subsidiary	Profit
Calgary	$59,426
Montreal	$515,244
Ottawa	$111,792
Toronto	$310,045
Vancouver	$3,129,811

Region=Central America/Caribbean

Subsidiary	Profit
Kingston	$2,159,259
Managua	$400,495
Mexico City	$355,763
San Juan	$615,338

Region=Eastern Europe

Subsidiary	Profit
Budapest	$394,378
Moscow	$542,541
Prague	$613,354
Warsaw	$757,966

Region=Middle East

Subsidiary	Profit
Al-Khobar	$1,117,382
Dubai	$1,846,111
Tel Aviv	$2,461,406

Region=Pacific

Subsidiary	Profit
Auckland	$119,896
Canberra	$150,232
Jakarta	$628,254
Kuala Lumpur	$359,845
Manila	$825,942
Singapore	$135,496

Region=South America

Subsidiary	Profit
Bogota	$196,141
Buenos Aires	$113,240
Caracas	$761,381
La Paz	$507,942
Montevideo	$302,621
Santiago	$99,759
Sao Paulo	$350,848

Region=United States

Subsidiary	Profit
Chicago	$1,508,062
Los Angeles	$712,000
Minneapolis	$1,062,864
New York	$1,442,506
Seattle	$591,052

Region=Western Europe

Subsidiary	Profit
Copenhagen	$668,593
Geneva	$432,920
Heidelberg	$931,701
Lisbon	$867,914
London	$734,599
Madrid	$193,308
Paris	$602,553
Rome	$271,657

```
*Essential Reviews 3;
data date_example;
        set cr.dates;
        FormattedDate=RawDate;
        format formatteddate date9.;
run;
```

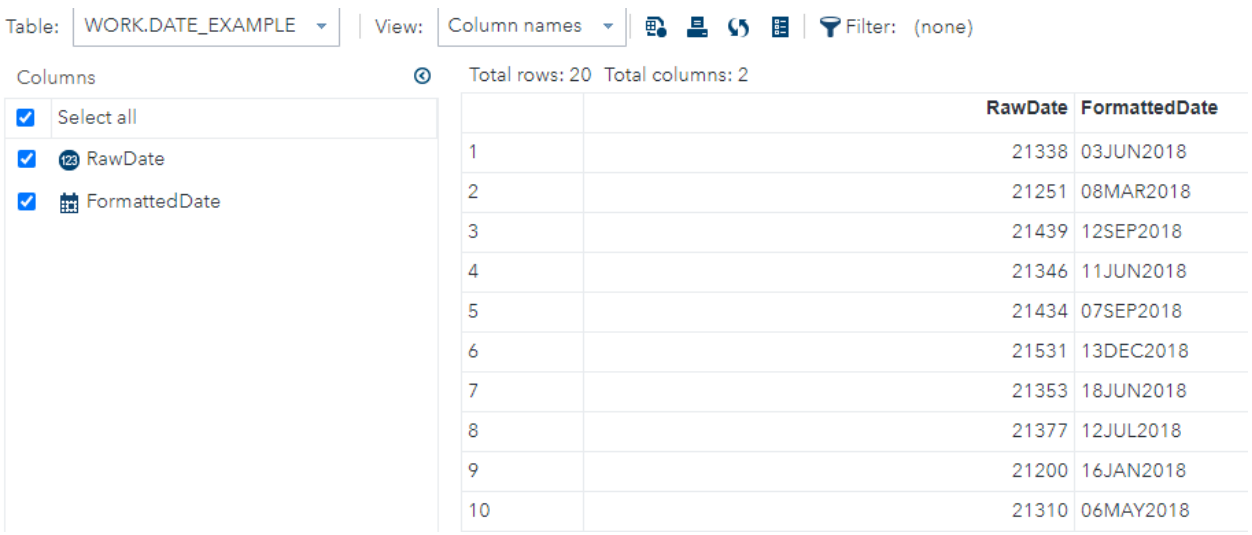


Table: WORK.DATE_EXAMPLE View: Column names Filter: (none)

Columns

Select all

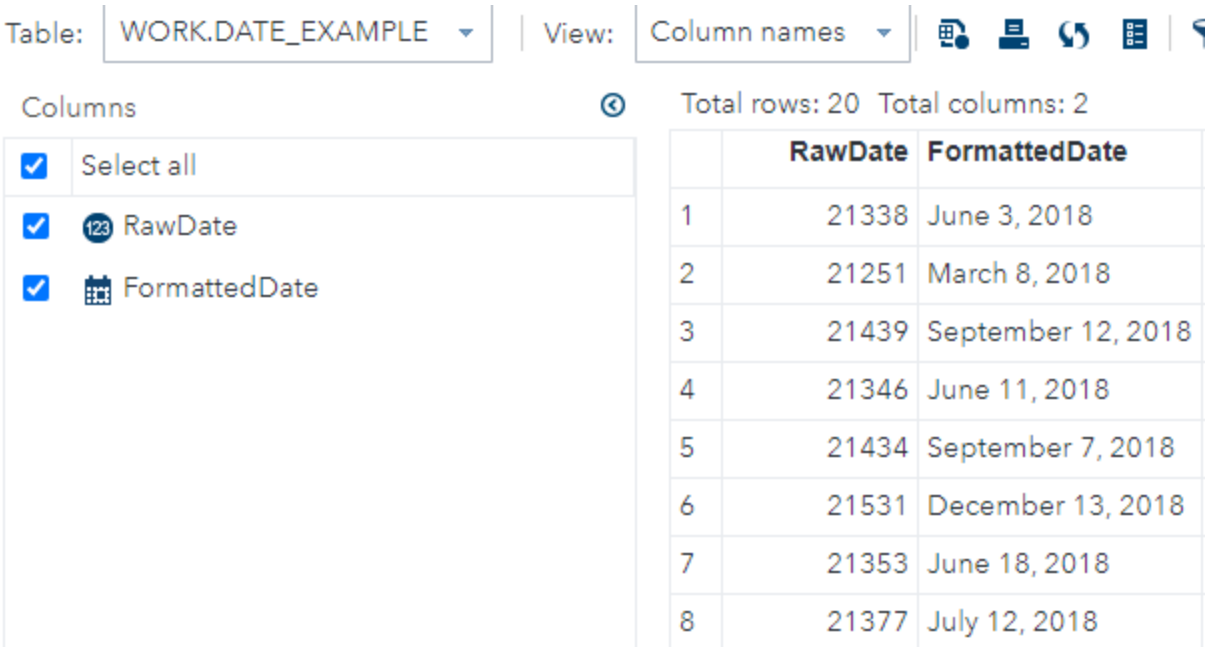
RawDate

FormattedDate

Total rows: 20 Total columns: 2

	RawDate	FormattedDate
1	21338	03JUN2018
2	21251	08MAR2018
3	21439	12SEP2018
4	21346	11JUN2018
5	21434	07SEP2018
6	21531	13DEC2018
7	21353	18JUN2018
8	21377	12JUL2018
9	21200	16JAN2018
10	21310	06MAY2018

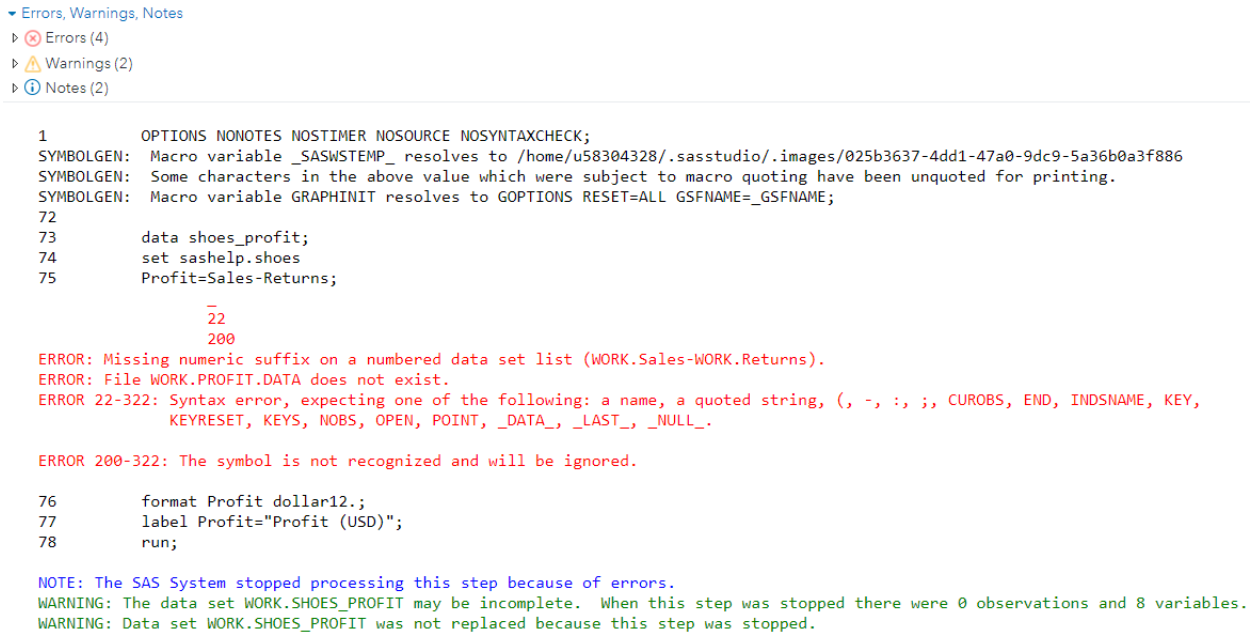
```
data date_example;
        set cr.dates;
        FormattedDate=RawDate;
        format formatteddate worddate.;
run;
```



	RawDate	FormattedDate
1	21338	June 3, 2018
2	21251	March 8, 2018
3	21439	September 12, 2018
4	21346	June 11, 2018
5	21434	September 7, 2018
6	21531	December 13, 2018
7	21353	June 18, 2018
8	21377	July 12, 2018

*sample of code missing ; in the set statement;

```
data shoes_profit;
        set sashelp.shoes
        Profit=Sales-Returns;
        format Profit dollar12.;
        label Profit="Profit (USD)";
run;
```



*Sample of code missing . after dollar12;

```
data shoes_profit;
        set sashelp.shoes;
        Profit=Sales-Returns;
        format Profit dollar12;
        label Profit="Profit (USD)";
run;
```

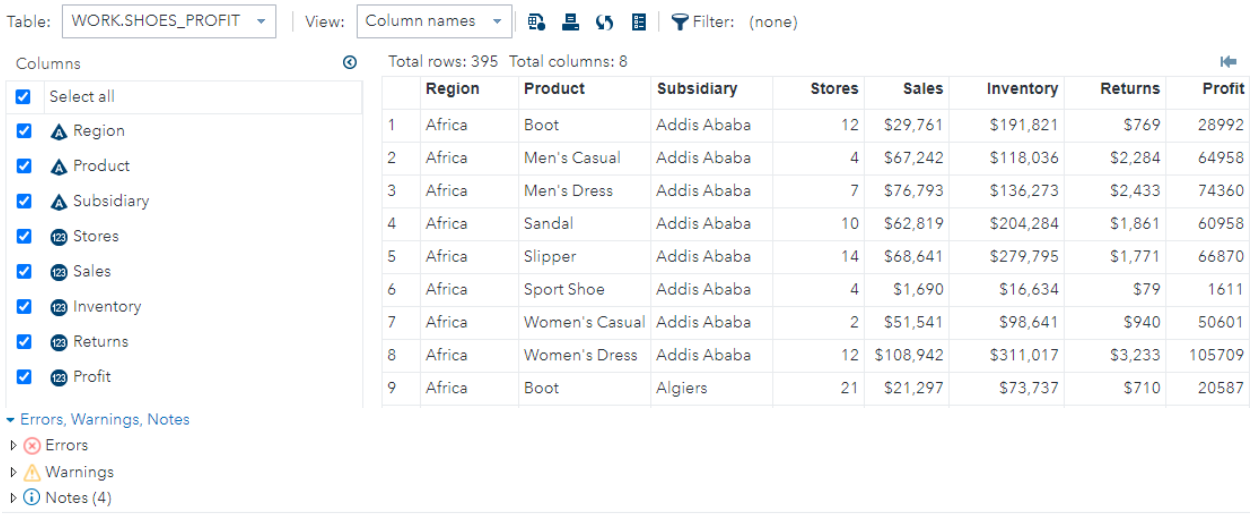


Table: WORK.SHOES_PROFIT | View: Column names | Filter: (none)

Columns

- Select all
- Region
- Product
- Subsidiary
- Stores
- Sales
- Inventory
- Returns
- Profit

Total rows: 395 Total columns: 8

	Region	Product	Subsidiary	Stores	Sales	Inventory	Returns	Profit
1	Africa	Boot	Addis Ababa	12	$29,761	$191,821	$769	28992
2	Africa	Men's Casual	Addis Ababa	4	$67,242	$118,036	$2,284	64958
3	Africa	Men's Dress	Addis Ababa	7	$76,793	$136,273	$2,433	74360
4	Africa	Sandal	Addis Ababa	10	$62,819	$204,284	$1,861	60958
5	Africa	Slipper	Addis Ababa	14	$68,641	$279,795	$1,771	66870
6	Africa	Sport Shoe	Addis Ababa	4	$1,690	$16,634	$79	1611
7	Africa	Women's Casual	Addis Ababa	2	$51,541	$98,641	$940	50601
8	Africa	Women's Dress	Addis Ababa	12	$108,942	$311,017	$3,233	105709
9	Africa	Boot	Algiers	21	$21,297	$73,737	$710	20587

Errors, Warnings, Notes

- Errors
- Warnings
- Notes (4)

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
SYMBOLGEN:  Macro variable _SASWSTEMP_ resolves to /home/u58304328/.sasstudio/.images/aa7c7f3e-99ba-4f28-8c0a-1a3bbd80d50a
SYMBOLGEN:  Some characters in the above value which were subject to macro quoting have been unquoted for printing.
SYMBOLGEN:  Macro variable GRAPHINIT resolves to GOPTIONS RESET=ALL GSFNAME=_GSFNAME;
72
73         data shoes_profit;
74         set sashelp.shoes;
75         Profit=Sales-Returns;
76         format Profit dollar12;
77         label Profit="Profit (USD)";
78         run;

NOTE: Variable dollar12 is uninitialized.
NOTE: There were 395 observations read from the data set SASHELP.SHOES.
NOTE: The data set WORK.SHOES_PROFIT has 395 observations and 8 variables.
```

```
data CanadaSales;
        set sashelp.prdsale;
        Diff=Actual-Predict;
        where Country='CANADA' and Quarter=1;
run;
```

*Original;

```
data=CanadaSales;
        set sashelp.prdsale;
        Actual-Predict=Diff;
        where Country=CANADA and Quarter=1;
run;
```

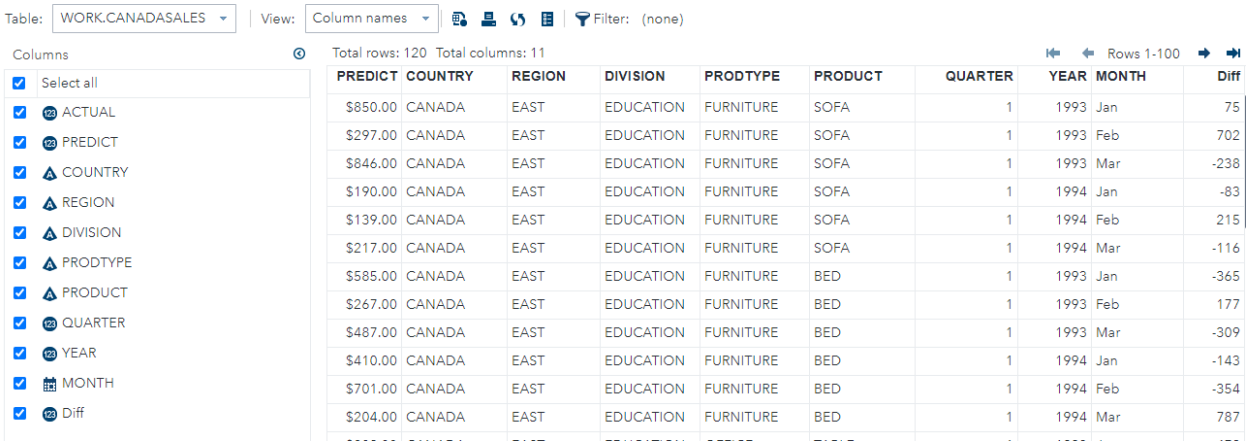


Table: WORK.CANADASALES | View: Column names | Filter: (none)

Columns: Select all, ACTUAL, PREDICT, COUNTRY, REGION, DIVISION, PRODTYPE, PRODUCT, QUARTER, YEAR, MONTH, Diff

Total rows: 120 Total columns: 11 — Rows 1-100

PREDICT	COUNTRY	REGION	DIVISION	PRODTYPE	PRODUCT	QUARTER	YEAR	MONTH	Diff
$850.00	CANADA	EAST	EDUCATION	FURNITURE	SOFA	1	1993	Jan	75
$297.00	CANADA	EAST	EDUCATION	FURNITURE	SOFA	1	1993	Feb	702
$846.00	CANADA	EAST	EDUCATION	FURNITURE	SOFA	1	1993	Mar	-238
$190.00	CANADA	EAST	EDUCATION	FURNITURE	SOFA	1	1994	Jan	-83
$139.00	CANADA	EAST	EDUCATION	FURNITURE	SOFA	1	1994	Feb	215
$217.00	CANADA	EAST	EDUCATION	FURNITURE	SOFA	1	1994	Mar	-116
$585.00	CANADA	EAST	EDUCATION	FURNITURE	BED	1	1993	Jan	-365
$267.00	CANADA	EAST	EDUCATION	FURNITURE	BED	1	1993	Feb	177
$487.00	CANADA	EAST	EDUCATION	FURNITURE	BED	1	1993	Mar	-309
$410.00	CANADA	EAST	EDUCATION	FURNITURE	BED	1	1994	Jan	-143
$701.00	CANADA	EAST	EDUCATION	FURNITURE	BED	1	1994	Feb	-354
$204.00	CANADA	EAST	EDUCATION	FURNITURE	BED	1	1994	Mar	787

```
/*This program analyzes blood pressure
in the SASHELP.HEART table.*/

proc freq data=sashelp.heart;
   table BP_Status;
run;

proc means data=sashelp.heart min max mean maxdec=0;
   var systolic diastolic;
   class BP_Status;
run;

*Original;
*This program analyzes blood pressure
in the SASHELP.HEART table.*/
```

```
proc freq data=sashelp.heart
   table BP_Status;
run;

proc means data=sashelp.heart min max average maxdec=0;
   var systolic diastolic;
   class BP_Status;
run;
```

The FREQ Procedure

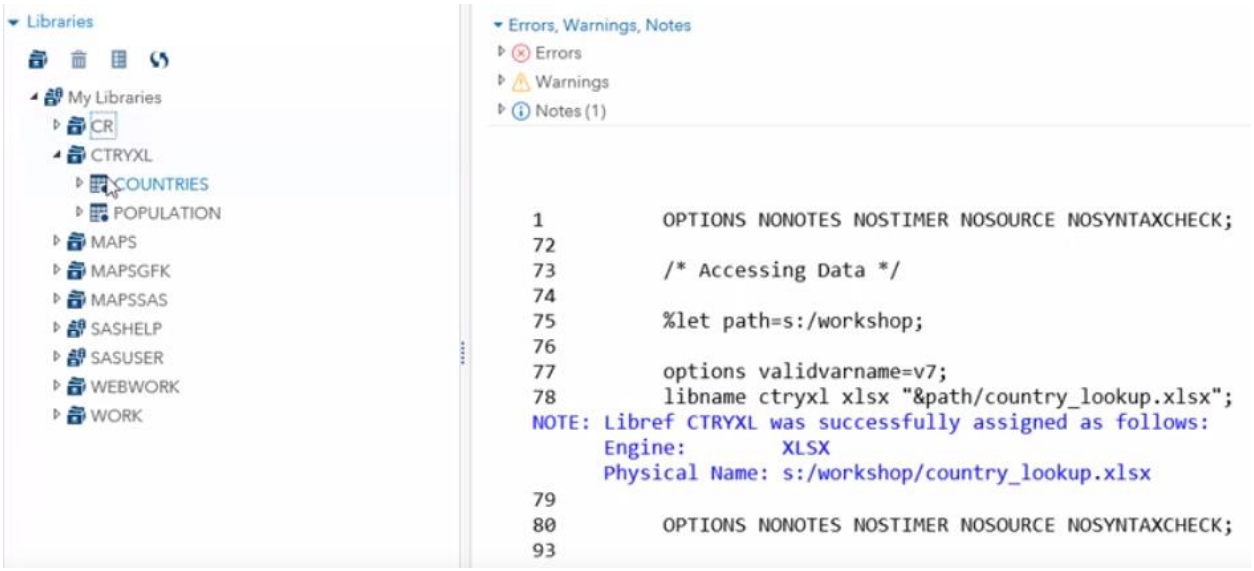
Blood Pressure Status				
BP_Status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
High	2267	43.52	2267	43.52
Normal	2143	41.14	4410	84.66
Optimal	799	15.34	5209	100.00

The MEANS Procedure

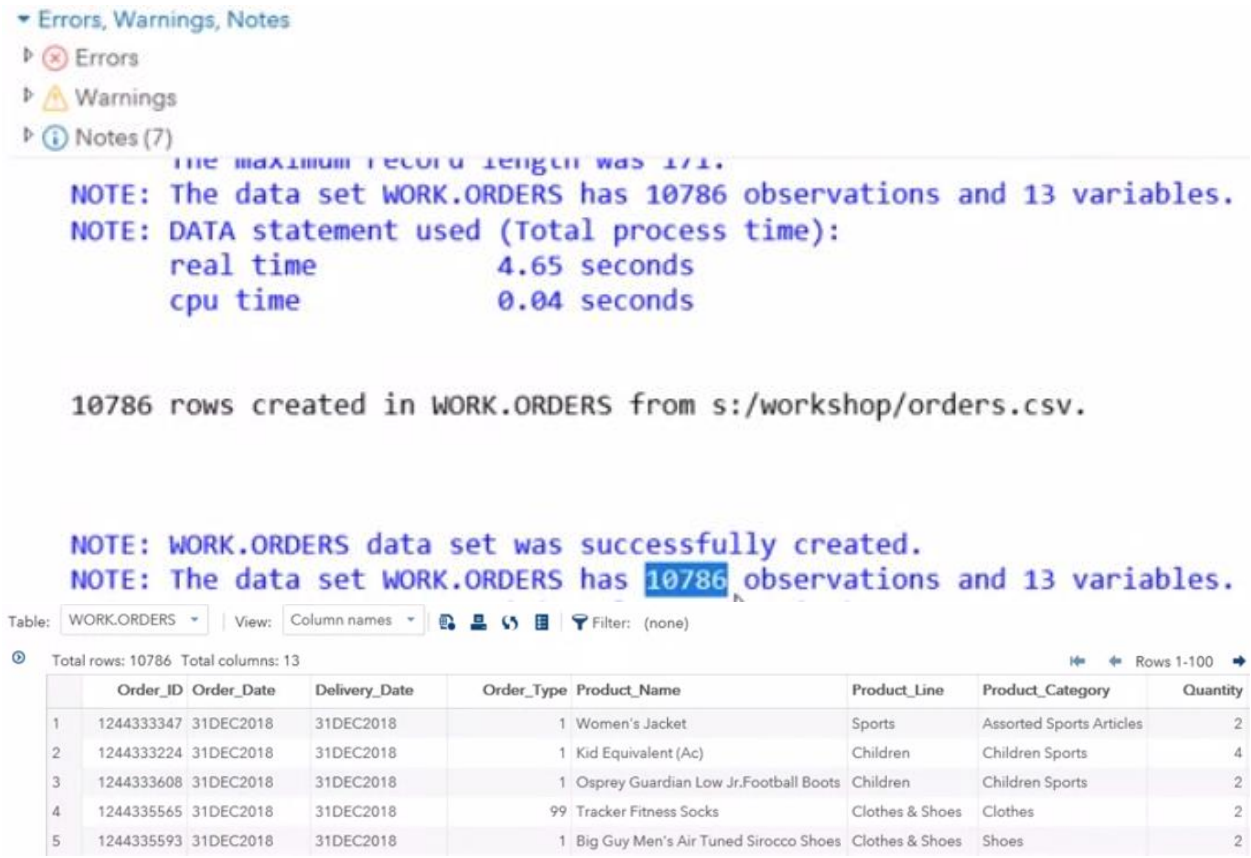
Blood Pressure Status	N Obs	Variable	Minimum	Maximum	Mean
High	2267	Systolic Diastolic	112 52	300 160	156 96
Normal	2143	Systolic Diastolic	101 54	140 88	127 80
Optimal	799	Systolic Diastolic	82 50	118 78	110 70

Accessing Data Review

```
/* Accessing Data  */

%let path=s:/workshop;

options validvarname=v7;
libname ctryxl xlsx "&path/country_lookup.xlsx";
```




```
proc import datafile="&path/orders.csv" out=orders dbms=csv replace;
run;
*practice2;
libname cr "&path/output";
proc import datafile="&path/orders.csv" out=orders dbms=csv replace;
run;
```



	Order_ID	Order_Date	Delivery_Date	Order_Type	Product_Name	Product_Line	Product_Category	Quantity
1	1244333347	31DEC2018	31DEC2018	1	Women's Jacket	Sports	Assorted Sports Articles	2
2	1244333224	31DEC2018	31DEC2018	1	Kid Equivalent (Ac)	Children	Children Sports	4
3	1244333608	31DEC2018	31DEC2018	1	Osprey Guardian Low Jr.Football Boots	Children	Children Sports	2
4	1244335565	31DEC2018	31DEC2018	99	Tracker Fitness Socks	Clothes & Shoes	Clothes	2
5	1244335593	31DEC2018	31DEC2018	1	Big Guy Men's Air Tuned Sirocco Shoes	Clothes & Shoes	Shoes	2

```
proc contents data=cr.orders;

run;


proc contents data=ctryxl.countries;

run;
```

The CONTENTS Procedure

Data Set Name	CR.ORDERS	Observations	10786
Member Type	DATA	Variables	13
Engine	V9	Indexes	0
Created	10/16/2018 12:24:45	Observation Length	176
Last Modified	10/16/2018 12:24:45	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

Engine/Host Dependent Information	
Data Set Page Size	65536
Number of Data Set Pages	30
First Data Page	1
Max Obs per Page	371
Obs in First Data Page	358
Number of Data Set Repairs	0
ExtendObsCounter	YES
Filename	s:\workshop\output\orders.sas7bdat
Release Created	9.0401M5
Host Created	X64_10PRO
Owner Name	CARYNT\stever
File Size	2MB
File Size (bytes)	2031616

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
10	Cost_Price	Num	8	BEST12.	BEST32.
13	Customer_Age_Group	Char	13	$13.	$13.
12	Customer_Continent	Char	19	$19.	$19.
11	Customer_Country	Char	4	$4.	$4.
3	Delivery_Date	Num	8	DATE9.	DATE9.
2	Order_Date	Num	8	DATE9.	DATE9.
1	Order_ID	Num	8	BEST12.	BEST32.
4	Order_Type	Num	8	BEST12.	BEST32.
7	Product_Category	Char	26	$26.	$26.
6	Product_Line	Char	17	$17.	$17.
5	Product_Name	Char	39	$39.	$39.
8	Quantity	Num	8	BEST12.	BEST32.
9	Retail_Price	Num	8	BEST12.	BEST32.

The CONTENTS Procedure

Data Set Name	CTRYXL.countries	Observations	.
Member Type	DATA	Variables	4
Engine	XLSX	Indexes	0
Created	.	Observation Length	0
Last Modified	.	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	Default		
Encoding	Default		

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
1	Country_Key	Char	2	$2.	$2.	Country_Key
4	Country_Name	Char	24	$24.	$24.	Country_Name
2	Lat	Num	8	BEST.		Lat
3	Lon	Num	8	BEST.		Lon

```
*print all worksheets;
proc contents data=ctryxl._all_;
run;
```

The CONTENTS Procedure

Directory	
Libref	CTRYXL
Engine	XLSX
Physical Name	s:/workshop/country_lookup.xlsx
Por	

#	Name	Member Type	DBMSTYPE
1	COUNTRIES	DATA	TABLE
2	POPULATION	DATA	TABLE

The CONTENTS Procedure

Data Set Name	CTRYXL.COUNTRIES	Observations	.
Member Type	DATA	Variables	4
Engine	XLSX	Indexes	0
Created	.	Observation Length	0
Last Modified	.	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	Default		
Encoding	Default		

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
1	Country_Key	Char	2	$2.	$2.	Country_Key
4	Country_Name	Char	24	$24.	$24.	Country_Name
2	Lat	Num	8	BEST.		Lat
3	Lon	Num	8	BEST.		Lon

The CONTENTS Procedure

Data Set Name	CTRYXL.POPULATION	Observations	.
Member Type	DATA	Variables	3
Engine	XLSX	Indexes	0
Created	.	Observation Length	0
Last Modified	.	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	Default		
Encoding	Default		

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
1	Country_Name	Char	24	$24.	$24.	Country_Name
3	Population	Num	8	COMMA15.		Population
2	Region	Char	4	$4.	$4.	Region

*print all worksheets as tables only without details;

proc contents data=ctryxl._all_ nods;

run;

The CONTENTS Procedure

Directory	
Libref	CTRYXL
Engine	XLSX
Physical Name	s:/workshop/country_lookup.xlsx
Por	.

#	Name	Member Type	DBMSTYPE
1	COUNTRIES	DATA	TABLE
2	POPULATION	DATA	TABLE

*print all cr library tables only;

proc contents data=cr._all_ nods;

run;

The CONTENTS Procedure

Directory	
Libref	CR
Engine	V9
Physical Name	/home/u58304328/ECRB94/data
Filename	/home/u58304328/ECRB94/data
Inode Number	12962730013
Access Permission	rwxr-xr-x
Owner Name	u58304328
File Size	4KB
File Size (bytes)	4096

#	Name	Member Type	File Size	Last Modified
1	CONTINENT_CODES	DATA	256KB	04/28/2021 06:12:06
2	COUNTRY_CLEAN	DATA	256KB	04/28/2021 06:12:06
3	COUNTRY_INFO	DATA	256KB	04/28/2021 06:12:06
4	DATES	DATA	256KB	04/28/2021 06:12:06
5	DEMOGRAPHICS	DATA	256KB	04/28/2021 06:12:06
6	EMPLOYEE	DATA	256KB	04/28/2021 06:12:06
7	EMPLOYEE_ADDRESSES	DATA	256KB	04/28/2021 06:12:06
8	EMPLOYEE_CURRENT	DATA	256KB	04/28/2021 06:12:06
9	EMPLOYEE_DONATIONS	DATA	256KB	04/28/2021 06:12:06
10	EMPLOYEE_NEW	DATA	256KB	04/28/2021 06:12:06
11	EMPLOYEE_RAW	DATA	256KB	04/28/2021 06:12:06
12	EMPLOYEE_TRAINING	DATA	256KB	04/28/2021 06:12:06
13	EMP_SERVICE	DATA	256KB	04/28/2021 06:12:06
14	M7_SALES	DATA	256KB	04/28/2021 06:12:06
15	M8_SALES	DATA	256KB	04/28/2021 06:12:06
16	M9_SALES	DATA	256KB	04/28/2021 06:12:06
17	ORDERS	DATA	2MB	04/28/2021 06:12:06
18	ORDERS2017	DATA	1MB	04/28/2021 06:12:06
19	PRODUCTS	DATA	384KB	04/28/2021 06:12:06
20	PROFIT	DATA	2MB	04/28/2021 06:12:06
21	PROFIT_SUMMARY	DATA	256KB	04/28/2021 06:12:06
22	QTR_SALES	DATA	256KB	04/28/2021 06:12:06
23	SALES	DATA	256KB	04/28/2021 06:12:06
24	SHOES_SUMMARY	DATA	256KB	04/28/2021 06:12:06
25	TOURISM	DATA	768KB	04/28/2021 06:12:06

```
*Quiz1 answer;
proc import datafile="&path/data/payroll.csv" out=payroll dbms=csv replace;
   guessingrows=max;
run;
```

proc contents data=payroll;

run;

The CONTENTS Procedure

Data Set Name	WORK.PAYROLL	Observations	424
Member Type	DATA	Variables	7
Engine	V9	Indexes	0
Created	04/28/2021 22:58:23	Observation Length	56
Last Modified	04/28/2021 22:58:23	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	2334
Obs in First Data Page	424
Number of Data Set Repairs	0
Filename	/saswork/SAS_work57A60000D66C_odaws04-usw2.oda.sas.com/SAS_workB1550000D66C_odaws04-usw2.oda.sas.com/payroll.sas7bdat
Release Created	9.0401M6
Host Created	Linux
Inode Number	536970666
Access Permission	rw-r--r--
Owner Name	u58304328
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
3	Birth_Date	Num	8	DATE7.	DATE7.
7	Dependents	Num	8	BEST12.	BEST32.
4	Employee_Hire_Date	Num	8	DATE7.	DATE7.
1	Employee_ID	Num	8	BEST12.	BEST32.
5	Employee_Term_Date	Num	8	DATE7.	DATE7.
6	Marital_Status	Char	1	$1.	$1.
2	Salary	Num	8	BEST12.	BEST32.

Table: WORK.PAYROLL | View: Column names | Filter: (none)

Columns

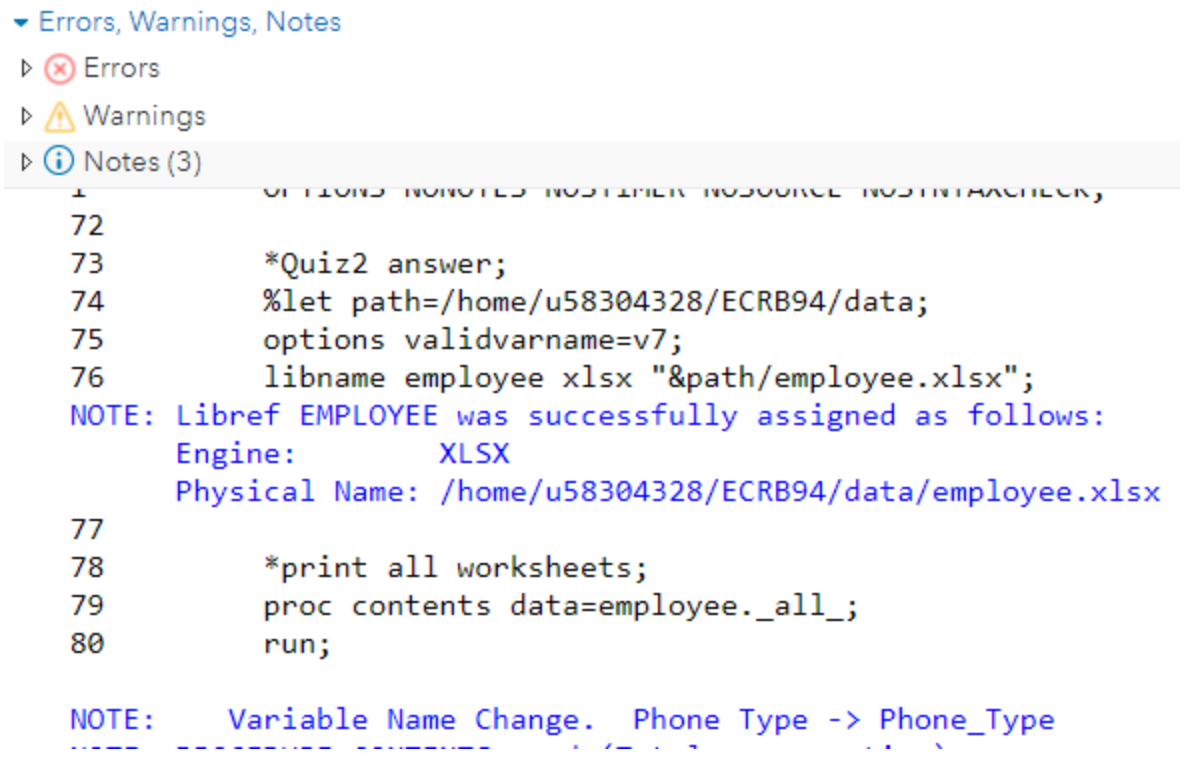
- Select all
- Employee_ID
- Salary
- Birth_Date
- Employee_Hire_Date
- Employee_Term_Date
- Marital_Status
- Dependents

Property	Value
Label	
Name	
Length	
Type	
Format	
Informat	

Total rows: 424 Total columns: 7 Rows

	Employee_ID	Salary	Birth_Date	Employee_Hire_Date	Employee_Term_Date	Marital_Status
1	120101	163040	18AUG81	01JUL14	.	S
2	120102	108255	11AUG74	01JUN00	.	O
3	120103	87975	22JAN54	01JAN85	.	M
4	120104	46230	11MAY59	01JAN92	.	M
5	120105	27110	21DEC79	01MAY10	.	S
6	120106	26960	23DEC49	01JAN85	.	M
7	120107	30475	21JAN54	01FEB85	.	M
8	120108	27660	23FEB89	01AUG17	.	S
9	120109	26495	15DEC91	01OCT17	.	M
10	120110	28615	20NOV54	01NOV90	.	M
11	120111	26895	23JUL54	01NOV85	.	M
12	120112	26550	17FEB74	01JUL01	.	M
13	120113	26870	10MAY49	01JAN85	.	S
14	120114	31285	08FEB49	01JAN85	.	M
15	120115	26500	08MAY89	01AUG16	.	M
16	120116	29250	13JUN64	01FEB91	.	S
17	120117	31670	11SEP69	01APR97	.	O
18	120118	28090	03JUN64	01JUL95	.	M
19	120119	30255	21DEC74	01JAN09	.	M
20	120120	27645	05MAY49	01JAN85	.	M
21	120121	26600	02AUG49	01JAN85	.	M
22	120122	27475	27JUL59	01JUL89	.	S
23	120123	26190	28SEP69	01OCT96	31JAN16	M

```
*Quiz2 answer;

%let path=/home/u58304328/ECRB94/data;

options validvarname=v7;

libname employee xlsx "&path/employee.xlsx";


*print all worksheets;

proc contents data=employee._all_;

run;
```


Errors, Warnings, Notes
Errors
Warnings
Notes (3)
72
73 *Quiz2 answer;
74 %let path=/home/u58304328/ECRB94/data;
75 options validvarname=v7;
76 libname employee xlsx "&path/employee.xlsx";
NOTE: Libref EMPLOYEE was successfully assigned as follows:
Engine: XLSX
Physical Name: /home/u58304328/ECRB94/data/employee.xlsx
77
78 *print all worksheets;
79 proc contents data=employee._all_;
80 run;
NOTE: Variable Name Change. Phone Type -> Phone_Type

The CONTENTS Procedure

Directory	
Libref	EMPLOYEE
Engine	XLSX
Physical Name	/home/u58304328/ECRB94/data/employee.xlsx
Por	.

#	Name	Member Type	DBMSTYPE
1	ADDRESSES	DATA	TABLE
2	PHONES	DATA	TABLE

The CONTENTS Procedure

Data Set Name	EMPLOYEE.ADDRESSES	Observations	.
Member Type	DATA	Variables	9
Engine	XLSX	Indexes	0
Created	.	Observation Length	0
Last Modified	.	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	Default		
Encoding	Default		

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
6	City	Char	12	$12.	$12.	City
9	Country	Char	2	$2.	$2.	Country
1	Employee_ID	Num	8	BEST.		Employee_ID
2	Employee_Name	Char	28	$28.	$28.	Employee_Name
8	Postal_Code	Num	8	BEST.		Postal_Code
7	State	Char	2	$2.	$2.	State
3	Street_ID	Num	8	BEST.		Street_ID
5	Street_Name	Char	29	$29.	$29.	Street_Name
4	Street_Number	Num	8	BEST.		Street_Number

/* Accessing Data */

```
%let path=/home/u58304328/ECRB94;

options validvarname=v7;
libname ctryxl xlsx "&path/data/country_lookup.xlsx";
```

```
libname cr "&path/output";
proc import datafile="&path/data/orders.csv" out=cr.orders dbms=csv replace;
run;

proc contents data=cr.orders;
run;

proc contents data=ctryxl.countries;
run;
```

/* Exploring Data */

```
/* Validate Country Lookup Excel Table */

proc print data=ctryxl.countries(obs=30);
run;
```

Obs	Country_Key	Lat	Lon	Country_Name
1	AD	42.546245	1.601554	Andorra
2	AE	23.424076	53.847818	United Arab Emirates
3	AF	33.93911	67.709953	Afghanistan
4	AG	17.060816	-61.796428	Antigua/Barbuda
5	AG	17.060816	-61.796428	Antigua/Barbuda
6	AI	18.220554	-63.068615	Anguilla
7	AL	41.153332	20.168331	Albania
8	AM	40.069099	45.038189	Armenia
9	AN	12.226079	-69.060087	Netherlands Antilles
10	AO	-11.202692	17.873887	Angola

```
proc freq data=ctryxl.countries order=freq;
        tables Country_Key Country_Name;
run;
```

The FREQ Procedure

Country_Key

Country_Key	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AG	2	0.84	2	0.84
CF	2	0.84	4	1.69
GB	2	0.84	6	2.53
US	2	0.84	8	3.38
AD	1	0.42	9	3.80
AE	1	0.42	10	4.22
AF	1	0.42	11	4.64
AI	1	0.42	12	5.06

Country_Name

Country_Name	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Antigua/Barbuda	2	0.84	2	0.84
Afghanistan	1	0.42	3	1.27
Albania	1	0.42	4	1.69

```
proc print data=ctryxl.countries;
        where Country_Key in ('AG','CF','GB','US');
run;
```

Obs	Country_Key	Lat	Lon	Country_Name
1	AG	17.060816	-61.796428	Antigua/Barbuda
2	AG	17.060816	-61.796428	Antigua/Barbuda
3	CF	6.611111	20.939444	Central African Rep.
4	CF	6.611111	20.939444	Central African Republic
5	GB	55.378051	-3.435973	United Kingdom
6	GB	55.378051	-3.435973	Great Britain
7	US	37.09024	-95.712891	United States
8	US	37.09024	-95.712891	United States of America

```
proc sort data=ctryxl.countries out=country_clean nodupkey dupout=dups;
        by Country_key;
run;
```

```
73          proc sort data=ctryxl.countries out=country_clean nodupkey dupout=dups;
74          by Country_key;
75          run;

NOTE: The import data set has 237 observations and 4 variables.
NOTE: There were 237 observations read from the data set CTRYXL.countries.
NOTE: 4 observations with duplicate key values were deleted.
NOTE: The data set WORK.COUNTRY_CLEAN has 233 observations and 4 variables.
NOTE: The data set WORK.DUPS has 4 observations and 4 variables.
```

```
proc print data=cr.orders;
        where Order_Date>Delivery_Date;
        var Order_ID Order_Date Delivery_Date;
run;
```

Obs	Order_ID	Order_Date	Delivery_Date
1050	1244098378	04DEC2018	01JAN1960
2473	1243695400	21OCT2018	01JAN1960
3288	1243440852	23SEP2018	01JAN1960
4454	1243061668	12AUG2018	01JAN1960
5963	1242634685	26JUN2018	01JAN1960

```
proc freq data=cr.orders;
        tables Order_Type Customer_Country Customer_Continent;
run;
```

The FREQ Procedure

Order_Type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	8337	77.29	8337	77.29
2	1090	10.11	9427	87.40
3	1356	12.57	10783	99.97
99	3	0.03	10786	100.00

Customer_Country	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AE	2	0.02	2	0.02
AT	12	0.11	14	0.13
AU	731	6.78	745	6.91
BE	416	3.86	1161	10.76
CA	44	0.41	1205	11.17
CH	13	0.12	1218	11.29
CZ	1	0.01	1219	11.30
DE	1282	11.89	2501	23.19
DK	182	1.69	2683	24.87
ES	1056	9.79	3739	34.67
FI	12	0.11	3751	34.78
FR	1281	11.88	5032	46.65
GB	1213	11.25	6245	57.90
GR	2	0.02	6247	57.92

HU	1	0.01	6249	57.94
IE	2	0.02	6251	57.95
IL	1	0.01	6252	57.96
IT	1383	12.82	7635	70.79
LU	3	0.03	7638	70.81
NL	704	6.53	8342	77.34
NO	8	0.07	8350	77.42
PT	12	0.11	8362	77.53
RU	1	0.01	8363	77.54
SE	9	0.08	8372	77.62
SI	2	0.02	8374	77.64
TR	8	0.07	8382	77.71
US	2386	22.12	10768	99.83
ZA	14	0.13	10782	99.96
au	1	0.01	10783	99.97
be	2	0.02	10785	99.99
fr	1	0.01	10786	100.00

Customer_Continent	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Africa	14	0.13	14	0.13
Asia	12	0.11	26	0.24
Australia/Pacific	732	6.79	758	7.03
Europe	7598	70.44	8356	77.47
North America	2430	22.53	10786	100.00

```
proc means data=cr.orders maxdec=0;
        var Quantity Retail_Price Cost_Price;
run;
```

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
Quantity	10786	2	1	-1	8
Retail_Price	10786	140	184	1	3191
Cost_Price	10786	65	89	0	1584

```
proc univariate data=cr.orders;
        var Quantity Retail_Price Cost_Price;
run;
```

Extreme Observations

Lowest		Highest	
Value	Obs	Value	Obs
-1	150	7	9396
-1	100	7	9670
-1	24	7	9964
1	10786	8	31
1	10785	8	617

The UNIVARIATE Procedure
Variable: Retail_Price

Moments			
N	10786	Sum Weights	10786
Mean	139.604041	Sum Observations	1505769.19
Std Deviation	183.543178	Variance	33688.0982
Skewness	5.03522958	Kurtosis	45.5031273
Uncorrected SS	573537603	Corrected SS	363326139
Coeff Variation	131.474115	Std Error Mean	1.76729089

```
proc sort data=sashelp.baseball out=baseball_Sort;
        by Team Name;
run;

proc print data=baseball_Sort;
        where Team in ("San Francisco", "Los Angeles", "Oakland");
        keep Name Team Salary Cr:;
run;

proc freq data=baseball_Sort order=freq;
        tables Position;
run;

proc means data=baseball_Sort maxdec=0;
        var Salary;
run;
```

▾ Errors, Warnings, Notes

▷ Errors

▷ Warnings

▷ Notes (10)

```
NOTE: There were 322 observations read from the data set SASHELP.BASEBALL.
NOTE: The data set WORK.BASEBALL_SORT has 322 observations and 24 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1086.15k
      OS Memory           38272.00k
      Timestamp           04/29/2021 07:05:08 AM
      Step Count                        180  Switch Count  2
      Page Faults                       0
      Page Reclaims                     130
      Page Swaps                        0
      Voluntary Context Switches        9
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           272


76
77         proc print data=baseball_Sort;
78         where Team in ("San Francisco", "Los Angeles", "Oakland");
79         keep Name Team Salary Cr:;
NOTE: The DROP and KEEP statements are not supported in procedure steps in this release of the SAS System. Therefore, these
      statements are ignored.
80         run;

NOTE: There were 40 observations read from the data set WORK.BASEBALL_SORT.
      WHERE Team in ('Los Angeles', 'Oakland', 'San Francisco');
```

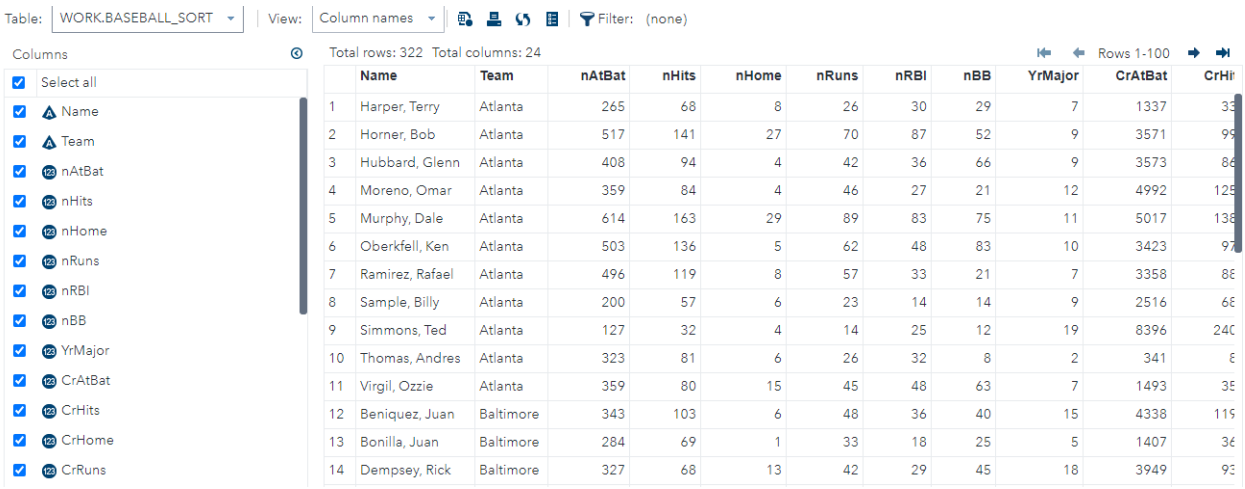
Obs	Name	Team	nAtBat	nHits	nHome	nRuns	nRBI	nBB	YrMajor	CrAtBat	CrHits	CrHome	CrRuns	CrRbi	CrBB	League	Division	Position	nOuts	nAssts	nError	Salary	Div	logSalary
135	Anderson, Dave	Los Angeles	216	53	1	31	15	22	4	926	210	9	118	69	114	National	West	3S	73	152	11	225.0	NW	5.41610
136	Brock, Greg	Los Angeles	325	76	16	33	52	37	5	1506	351	71	195	219	214	National	West	1B	726	87	3	385.0	NW	5.95324
137	Cabell, Enos	Los Angeles	277	71	2	27	29	14	15	5952	1647	60	753	596	259	National	West	1B	360	32	5	.	NW	.
138	Duncan, Mariano	Los Angeles	407	93	8	47	30	30	2	969	230	14	121	69	68	National	West	SS	172	317	25	150.0	NW	5.01064
139	Landreaux, Ken	Los Angeles	283	74	4	34	29	22	10	3919	1062	85	505	456	283	National	West	OF	145	5	7	737.5	NW	6.60327
140	Madlock, Bill	Los Angeles	379	106	10	38	60	30	14	6207	1906	146	859	803	571	National	West	3B	72	170	24	850.0	NW	6.74524
141	Marshall, Mike	Los Angeles	330	77	19	47	53	27	6	1928	516	90	247	288	161	National	West	RF	149	8	6	670.0	NW	6.50728
142	Matuszek, Len	Los Angeles	199	52	9	26	28	21	6	805	191	30	113	119	87	National	West	O1	235	22	5	265.0	NW	5.57973
143	Russell, Bill	Los Angeles	216	54	0	21	18	15	18	7318	1926	46	796	627	483	National	West	UT	103	84	5	.	NW	.
144	Sax, Steve	Los Angeles	633	210	6	91	56	59	6	3070	872	19	420	230	274	National	West	2B	367	432	16	90.0	NW	4.49981
145	Scioscia, Mike	Los Angeles	374	94	5	36	26	62	7	1968	519	26	181	199	288	National	West	C	756	64	15	875.0	NW	6.77422
146	Stubbs, Franklin	Los Angeles	420	95	23	55	58	37	3	646	139	31	77	77	61	National	West	LF	206	10	7	.	NW	.
147	Trevino, Alex	Los Angeles	202	53	4	31	26	27	9	1876	467	15	192	186	161	National	West	C	304	45	11	512.5	NW	6.23930

The FREQ Procedure

Position(s) in 1986				
Position	Frequency	Percent	Cumulative Frequency	Cumulative Percent
C	40	12.42	40	12.42
3B	32	9.94	72	22.36
1B	31	9.63	103	31.99
2B	31	9.63	134	41.61
OF	30	9.32	164	50.93
SS	30	9.32	194	60.25
CF	26	8.07	220	68.32
RF	26	8.07	246	76.40
LF	25	7.76	271	84.16
DH	16	4.97	287	89.13
UT	14	4.35	301	93.48
O1	4	1.24	305	94.72
3S	3	0.93	308	95.65
DO	2	0.62	310	96.27
OS	2	0.62	312	96.89
13	1	0.31	313	97.20
1O	1	0.31	314	97.52
23	1	0.31	315	97.83
2S	1	0.31	316	98.14
32	1	0.31	317	98.45
3O	1	0.31	318	98.76
CD	1	0.31	319	99.07
CS	1	0.31	320	99.38
OD	1	0.31	321	99.69
S3	1	0.31	322	100.00

The MEANS Procedure

Analysis Variable : Salary 1987 Salary in $ Thousands				
N	Mean	Std Dev	Minimum	Maximum
263	536	451	68	2460



	Name	Team	nAtBat	nHits	nHome	nRuns	nRBI	nBB	YrMajor	CrAtBat	CrHi
1	Harper, Terry	Atlanta	265	68	8	26	30	29	7	1337	33
2	Horner, Bob	Atlanta	517	141	27	70	87	52	9	3571	99
3	Hubbard, Glenn	Atlanta	408	94	4	42	36	66	9	3573	86
4	Moreno, Omar	Atlanta	359	84	4	46	27	21	12	4992	125
5	Murphy, Dale	Atlanta	614	163	29	89	83	75	11	5017	136
6	Oberkfell, Ken	Atlanta	503	136	5	62	48	83	10	3423	97
7	Ramirez, Rafael	Atlanta	496	119	8	57	33	21	7	3358	88
8	Sample, Billy	Atlanta	200	57	6	23	14	14	9	2516	68
9	Simmons, Ted	Atlanta	127	32	4	14	25	12	19	8396	240
10	Thomas, Andres	Atlanta	323	81	6	26	32	8	2	341	8
11	Virgil, Ozzie	Atlanta	359	80	15	45	48	63	7	1493	35
12	Beniquez, Juan	Baltimore	343	103	6	48	36	40	15	4338	115
13	Bonilla, Juan	Baltimore	284	69	1	33	18	25	5	1407	36
14	Dempsey, Rick	Baltimore	327	68	13	42	29	45	18	3949	93

*Answer p103 quiz 1.07;

```
proc freq data="&path/data/employee_raw.sas7bdat" order=freq nlevels ;
   tables EmpID Country Department;
run;


proc print data="&path/data/employee_raw.sas7bdat";
   where TermDate ne . and HireDate>TermDate;
   format salary dollar10. TermDate HireDate BirthDate date9.;
run;
```

The FREQ Procedure

Number of Variable Levels		
Variable	Label	Levels
EmpID		424
Country	Country	4
Department		17

EmpID	Frequency	Percent	Cumulative Frequency	Cumulative Percent
120107	2	0.47	2	0.47
120101	1	0.24	3	0.71
120102	1	0.24	4	0.94
120103	1	0.24	5	1.18

Country				
Country	Frequency	Percent	Cumulative Frequency	Cumulative Percent
US	311	73.18	311	73.18
AU	106	24.94	417	98.12
us	5	1.18	422	99.29
au	3	0.71	425	100.00

Department	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Sales	201	47.29	201	47.29
Administration	35	8.24	236	55.53
Stock & Shipping	26	6.12	262	61.65
IS	25	5.88	287	67.53
Marketing	20	4.71	307	72.24
Group HR Management	18	4.24	325	76.47
Purchasing	18	4.24	343	80.71
Accounts	17	4.00	360	84.71
Logistics Management	14	3.29	374	88.00
Concession Management	11	2.59	385	90.59
Sales Management	11	2.59	396	93.18
Accounts Management	9	2.12	405	95.29
Engineering	9	2.12	414	97.41
Executives	4	0.94	418	98.35
Group Financials	3	0.71	421	99.06
Secretary of the Board	2	0.47	423	99.53
Strategy	2	0.47	425	100.00

Obs	EmpID	Name	JobTitle	Department	ManagerID	Salary	Status	City	State	Country	BirthDate	HireDate	TermDate
45	121012	Broome, Carmelo	Service Assistant II	Administration	121009	$36,969	S	Miami-Dade	FL	US	26JAN1992	30NOV2015	01SEP2014

```
%let path=/home/u58304328/ECRB94;
```

```
options validvarname=v7;

proc contents data="&path/data/employee_raw.sas7bdat";
run;

/* Exploring Data */

/* Validate Employee_raw Table */

proc print data="&path/data/employee_raw.sas7bdat"(obs=30);
run;

proc freq data="&path/data/employee_raw.sas7bdat" order=freq;
        tables EmpID Department;
run;

proc print data="&path/data/employee_raw.sas7bdat";
        where Country NOT in ('AU','US');
run;

proc print data="&path/data/employee_raw.sas7bdat";
        where TermDate ne . and TermDate<HireDate;
        var EmpId TermDate HireDate;
run;

proc means data="&path/data/employee_raw.sas7bdat" maxdec=0;
        var Department;
run;
```

/*Question 1

If necessary, start SAS Studio and run libname.sas.

Write and submit a program to read the cr.employee_raw table and create a new sorted table named emp_sort.

Sort rows by all columns and remove entirely duplicated rows.

What is the value of EmpID for observation 10 in the emp_sort table?

*/

```
proc sort data=cr.employee_raw out=emp_sort noduprecs;
   by _all_;
run;

proc sort data=cr.employee_raw out=emp_sort nodupkey dupout=dups;
          by EmpID Name JobTitle Department ManagerID Salary Status City State Country BirthDate
HireDate TermDate;
run;
```

Table: WORK.EMP_SORT | View: Column names | Filter: (none)

Columns

- Select all
- EmpID
- Name
- JobTitle
- Department
- ManagerID
- Salary
- Status
- City
- State
- Country
- BirthDate
- HireDate
- TermDate

Total rows: 424 Total columns: 13 Rows 1-100

	EmpID	Name	JobTitle	Department	ManagerID	Salary	Status	City
1	120101	Lu, Patrick	Director	Sales Management	120261	$203,800.00	S	Sydney
2	120102	Zhou, Tom	Sales Manager	Sales Management	120101	$135,318.75	O	Melbourne
3	120103	Dawes, Wilson	Sales Manager	Sales Management	120101	$109,968.75	M	Sydney
4	120104	Billington, Kareen	Administration Manage	Administration	120101	$57,787.50	M	Sydney
5	120105	Povey, Liz	Secretary I	Administration	120101	$33,887.50	S	Melbourne
6	120106	Hornsey, John	Office Assistant II	Administration	120104	$33,700.00	M	Sydney
7	120107	Sheedy, Sherie	Office Assistant III	Administration	120104	$38,093.75	M	Melbourne
8	120108	Gromek, Gladys	Warehouse Assistant II	Administration	120104	$34,575.00	S	Melbourne
9	120109	Baker, Gabriele	Warehouse Assistant I	Administration	120104	$33,118.75	M	Sydney
10	120110	Entwisle, Dennis	Warehouse Assistant III	Administration	120104	$35,768.75	M	Sydney
11	120111	Spillane, Ubaldo	Security Guard II	Administration	120104	$33,618.75	M	Sydney
12	120112	Glattback, Ellis	Security Guard I	Administration	120104	$33,187.50	M	Melbourne
13	120113	Horsey, Riu	Security Guard II	Administration	120104	$33,587.50	S	Melbourne
14	120114	Buddery, Jeannette	Security Manager	Administration	120104	$39,106.25	M	Sydney

/* Validate Employee_raw Table */

/* Question 2 */

/* In your program, write and submit a procedure step to read the emp_sort table and

create a listing of all employees with a JobTitle that includes Logistics. */

/* What year was the most recent person hired in the group? Note: Type your answer as a 4-digit year. */

```
proc print data=emp_sort;

   where JobTitle like '%Logistics%';

   format salary dollar10. TermDate HireDate BirthDate date9.;

run;

proc print data=emp_sort;
        where JobTitle like "%Logistic%";
        format BirthDate HireDate TermDate date9.;
run;
```

Obs	EmpID	Name	JobTitle	Department	ManagerID	Salary	Status	City	State	Country	BirthDate	HireDate	TermDate
121	120656	Amos, Salley	Logistics Coordinator II	Logistics Management	120660	$53,212.50	O	San Diego	CA	US	28JAN1980	01MAR2010	.
122	120657	Weisbarth, Theresa	Logistics Coordinator I	Logistics Management	120660	$45,137.50	S	San Diego	CA	US	07JAN1970	01NOV2001	.
123	120658	Kennedy, Kenneth	Logistics Coordinator II	Logistics Management	120660	$53,106.25	O	San Diego	CA	US	20DEC1960	01FEB1992	.
125	120660	Smith, Robert	Logistics Manager	Logistics Management	120659	$76,406.25	S	Miami-Dade	FL	US	06JUN1984	01MAR2016	.
126	120661	Racine, Cynthia	Senior Logistics Manager	Logistics Management	120659	$106,868.75	M	San Diego	CA	US	27NOV1960	01JAN1995	30JUN2014
130	120665	Leacock, Jill	Senior Logistics Manager	Logistics Management	120659	$100,087.50	M	San Diego	CA	US	24OCT1980	01MAR2013	.
131	120666	Onuscheck, John	Logistics Manager	Logistics Management	120659	$80,693.75	S	San Diego	CA	US	22JUN1975	01DEC2002	30APR2015

```
/*Question 3

In your program, write and submit a procedure step to read the emp_sort table and answer this
question:

What is the average salary for employees with a hire date on or after January 1, 2010, and a missing
value for TermDate?

Note: Type your answer exactly as the value is displayed.

*/

proc means data=emp_sort n mean;

   where HireDate >= "01JAN2010"d and TermDate is missing;

   var salary;

run;
```

The MEANS Procedure

Analysis Variable : Salary	
N	Mean
116	40774.46

```
proc means data=emp_sort maxdec=2;

        var Salary;

        where Year(HireDate)>=2010 and TermDate=.;
```

run;

The MEANS Procedure

Analysis Variable : Salary				
N	Mean	Std Dev	Minimum	Maximum
116	40774.46	18832.27	31275.00	203800.00

/*Question 4

In your program, write and submit a procedure step to read the emp_sort table and answer this question:

What is the third highest salary among all employees? Note: Type your answer exactly the way the value is displayed.

*/

proc univariate data=emp_sort;

var Salary;

run;

The UNIVARIATE Procedure
Variable: Salary

Moments			
N	424	Sum Weights	424
Mean	46871.8573	Sum Observations	19873667.5
Std Deviation	34005.7266	Variance	1156389444
Skewness	6.15573422	Kurtosis	51.4134239
Uncorrected SS	1.42067E12	Corrected SS	4.89153E11
Coeff Variation	72.5504142	Std Error Mean	1651.46408

Basic Statistical Measures			
Location		Variability	
Mean	46871.86	Std Deviation	34006
Median	35856.25	Variance	1156389444
Mode	31275.00	Range	405413
		Interquartile Range	12078

Note: The mode displayed is the smallest of 3 modes with a count of 3.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	28.382	Pr > |t|	<.0001
Sign	M	212	Pr >= |M|	<.0001
Signed Rank	S	45050	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	433800.0
99%	207885.0
95%	81406.3
90%	68450.0
75% Q3	45503.1
50% Median	35856.3
25% Q1	33425.0
10%	32381.3
5%	31606.3
1%	30487.5
0% Min	28387.5

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
28387.5	360	207885	100
30018.8	91	243190	101
30031.3	96	243606	417
30125.0	90	268455	102
30487.5	408	433800	99