Preparing for the SAS Programming Certification

Week 2: Preparing Data Review, Analyzing Data Review, Exporting 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. */
/*********************************************/
/* Preparing Data */


data profit;
        set cr.orders;
        length Order_Source $ 8;
        where Delivery_Date>=Order_Date;
        Customer_Country=upcase(Customer_Country);
        if Quantity<0 then Quantity=.;
        Profit=(Retail_Price-Cost_Price)*Quantity;
        format Profit dollar12.2;
        Ship_Days=Delivery_Date-Order_Date;
        Age_Range=substr(Customer_Age_Group, 1, 5);
        if Order_Type=1 then Order_Source="Retail";
        else if Order_Type=2 then Order_Source="Phone";
        else if Order_Type=3 then Order_Source="Internet";
        else Order_Source="Unknown";
        drop Retail_Price Cost_Price Customer_Age_Group Order_Type;
run;
```

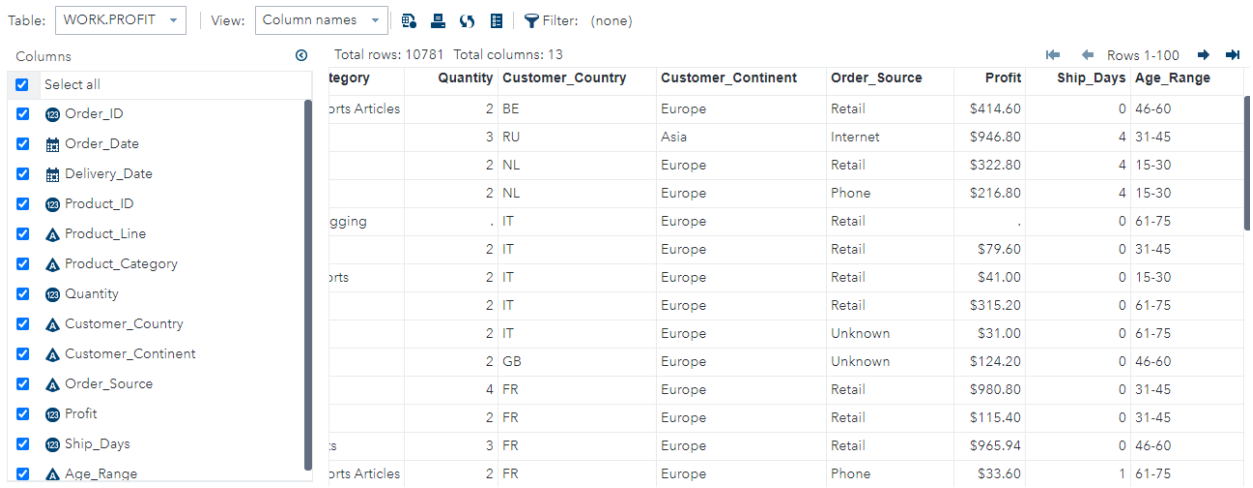


Table: WORK.PROFIT | View: Column names | Filter: (none)

Columns: Select all, Order_ID, Order_Date, Delivery_Date, Product_ID, Product_Line, Product_Category, Quantity, Customer_Country, Customer_Continent, Order_Source, Profit, Ship_Days, Age_Range

Total rows: 10781 Total columns: 13 — Rows 1-100

tegory	Quantity	Customer_Country	Customer_Continent	Order_Source	Profit	Ship_Days	Age_Range
orts Articles	2	BE	Europe	Retail	$414.60	0	46-60
	3	RU	Asia	Internet	$946.80	4	31-45
	2	NL	Europe	Retail	$322.80	4	15-30
	2	NL	Europe	Phone	$216.80	4	15-30
gging	.	IT	Europe	Retail	.	0	61-75
	2	IT	Europe	Retail	$79.60	0	31-45
orts	2	IT	Europe	Retail	$41.00	0	15-30
	2	IT	Europe	Retail	$315.20	0	61-75
	2	IT	Europe	Unknown	$31.00	0	61-75
	2	GB	Europe	Unknown	$124.20	0	46-60
	4	FR	Europe	Retail	$980.80	0	31-45
	2	FR	Europe	Retail	$115.40	0	31-45
s	3	FR	Europe	Retail	$965.94	0	46-60
orts Articles	2	FR	Europe	Phone	$33.60	1	61-75

```
proc sql;
        create table profit_country as
        select profit.*, Country_Name
        from profit inner join country_clean
        on profit.Customer_Country=country_clean.Country_Key
        order by Order_Date desc;
quit;
```

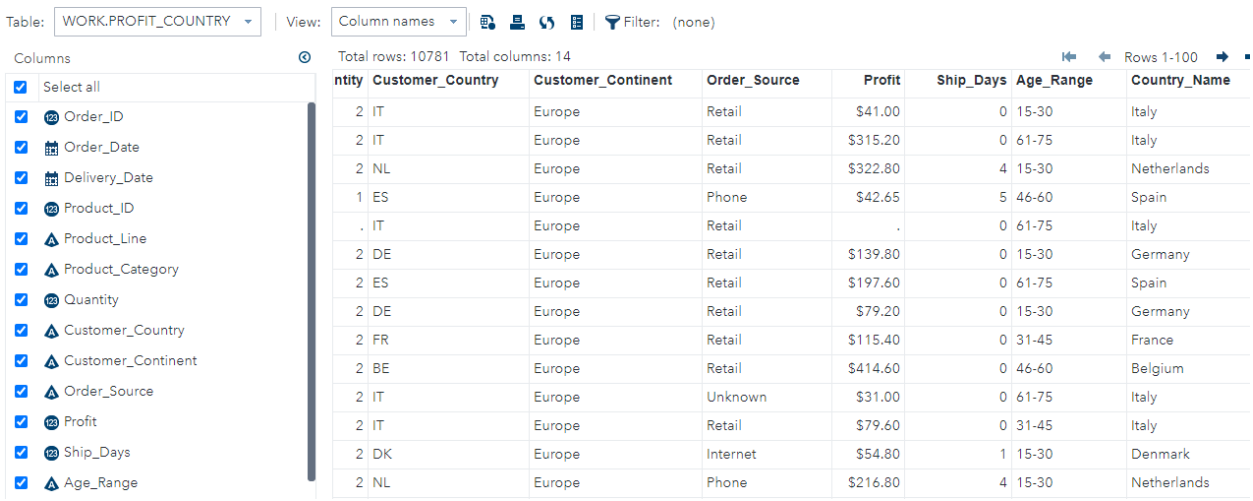


Table: WORK.PROFIT_COUNTRY | View: Column names | Filter: (none)

Columns: Select all, Order_ID, Order_Date, Delivery_Date, Product_ID, Product_Line, Product_Category, Quantity, Customer_Country, Customer_Continent, Order_Source, Profit, Ship_Days, Age_Range

Total rows: 10781 Total columns: 14 — Rows 1-100

ntity	Customer_Country	Customer_Continent	Order_Source	Profit	Ship_Days	Age_Range	Country_Name
2	IT	Europe	Retail	$41.00	0	15-30	Italy
2	IT	Europe	Retail	$315.20	0	61-75	Italy
2	NL	Europe	Retail	$322.80	4	15-30	Netherlands
1	ES	Europe	Phone	$42.65	5	46-60	Spain
.	IT	Europe	Retail	.	0	61-75	Italy
2	DE	Europe	Retail	$139.80	0	15-30	Germany
2	ES	Europe	Retail	$197.60	0	61-75	Spain
2	DE	Europe	Retail	$79.20	0	15-30	Germany
2	FR	Europe	Retail	$115.40	0	31-45	France
2	BE	Europe	Retail	$414.60	0	46-60	Belgium
2	IT	Europe	Unknown	$31.00	0	61-75	Italy
2	IT	Europe	Retail	$79.60	0	31-45	Italy
2	DK	Europe	Internet	$54.80	1	15-30	Denmark
2	NL	Europe	Phone	$216.80	4	15-30	Netherlands

```
data holiday2019;
        set sashelp.holiday;
        where end=. and rule=0;
        CountryCode=substr(Category,4,"_");
        Date=mdy(month, day, 2019);
```

```
        keep Desc CountryCode Date;
        *format Date date9.;
run;
*Answer;
data holiday2019;
  set sashelp.holiday;
        where end=. and rule=0;
        CountryCode=substr(category,4,2);
        Date=mdy(month, day, 2019);
        keep Desc CountryCode Date;
run;
```

Table: WORK.HOLIDAY2019 | View: Column names | Filter: (none)

Columns

- Select all
- desc
- CountryCode
- Date

Total rows: 10 Total columns: 3 Rows 1-10

	desc	CountryCode	Date
1	Boxing Day	CA	21909
2	Canadian Independence Day	CA	21731
3	Christmas	US	21908
4	Christmas	CA	21908
5	Halloween	US	21853
6	New Year's Day	US	21550
7	New Year's Day	CA	21550
8	U.S. Independence Day	US	21734
9	Valentine's Day	US	21594
10	Veterans Day	US	21864

/*Programming Question 2.02

If necessary, start SAS Studio and submit libname.sas. Write and submit a new program to do the following:

Create a new table named sales and read cr.employee. Include only employees in the Sales Department with no termination date (TermDate).

Create a new column named SalesLevel based on the following values of JobTitle:

JobTitle SalesLevel

Sales Rep. I Entry

Sales Rep. II or Sales Rep. III Middle

Sales Rep. IV Senior

Generate a report that includes the number of Sales employees in each level.

What is the total number of Sales employees?

What is the total number of middle-level sales reps?

*/

```
data sales;
        set cr.employee;
        length SalesLevel $ 6;
        where TermDate=. and Department="Sales";
        if JobTitle="Sales Rep. I" then SalesLevel="Entry";
        else if JobTitle in ("Sales Rep. II" "Sales Rep. III") then SalesLevel="Middle";
        else if JobTitle="Sales Rep. IV" then SalesLevel="Senior";
run;


proc freq data=sales;
        table SalesLevel;
run;
```

The FREQ Procedure

SalesLevel	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Entry	46	37.70	46	37.70
Middle	66	54.10	112	91.80
Senior	10	8.20	122	100.00

/*Programming Question 2.03

If necessary, start SAS Studio and submit libname.sas. Write and submit a program to do the following:

Create a new table named bonus and read cr.employee.

Exclude any employees with a known value for TermDate.

Use the YRDIF function to create a new column named YearsEmp that calculates the number of years that

each person has been employed as of 01JAN2019.

For employees that have been employed 10 years or more, create a new column named Bonus that is 3% of Salary.

Create another column named Vacation that is assigned the number 20.

For all other employees, calculate Bonus as 2% of Salary. Assign 15 as the value of Vacation.

Count the number of employees with 20 and 15 vacation days.

Sort the bonus table by YearsEmp in descending order.

How many employees are in the bonus table?

How many years has the employee in row number 1 of the sorted bonus table been employed?

How many employees have 20 vacation days?

What is the bonus amount for the last employee listed in the sorted bonus table?

*/

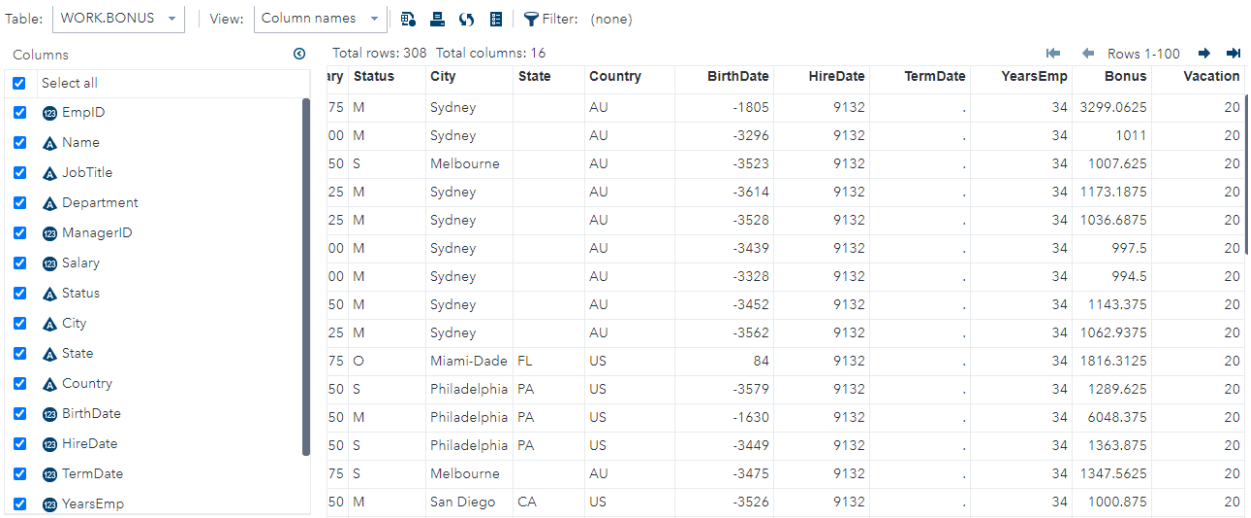
```
data bonus;
	set cr.employee;
	where TermDate=.;
	YearsEmp=yrdif(HireDate, "01JAN2019"d, "AGE");
	if YearsEmp>=10 then do;
		Bonus=Salary*0.03;
		Vacation=20;
	end;
	else do;
		Bonus=Salary*0.02;
		Vacation=15;
	end;
run;

proc freq data=bonus;
	table YearsEmp Vacation;
run;

proc sort data=bonus;
	by descending YearsEmp;
run;
```

The FREQ Procedure

YearsEmp	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	3	0.97	3	0.97
1.0849315068	5	1.62	8	2.60
1.1671232877	9	2.92	17	5.52
1.2520547945	8	2.60	25	8.12
1.3342465753	4	1.30	29	9.42
1.4191780822	7	2.27	36	11.69
1.504109589	6	1.95	42	13.64
1.5863013699	3	0.97	45	14.61
1.6712328767	6	1.95	51	16.56
1.7534246575	7	2.27	58	18.83
1.8383561644	8	2.60	66	21.43
1.9150684932	3	0.97	69	22.40
2.4191780822	2	0.65	71	23.05
2.7534246575	2	0.65	73	23.70
2.8383561644	2	0.65	75	24.35
2.9150684932	1	0.32	76	24.68
3.0849315068	1	0.32	77	25.00
3.3342465753	2	0.65	79	25.65
3.5863013699	1	0.32	80	25.97
3.6712328767	1	0.32	81	26.30
3.7534246575	1	0.32	82	26.62
3.9150684932	1	0.32	83	26.95
4.504109589	2	0.65	85	27.60
4.6712328767	1	0.32	86	27.92
4.8383561644	1	0.32	87	28.25
4.9150684932	2	0.65	89	28.90
5	2	0.65	91	29.55

30	1	0.32	251	81.49
30.084931507	1	0.32	252	81.82
30.252054795	1	0.32	253	82.14
30.334246575	2	0.65	255	82.79
31.084931507	1	0.32	256	83.12
31.167123288	1	0.32	257	83.44
31.753424658	3	0.97	260	84.42
32.334246575	1	0.32	261	84.74
32.419178082	1	0.32	262	85.06
32.671232877	1	0.32	263	85.39
33.167123288	1	0.32	264	85.71
33.252054795	1	0.32	265	86.04
33.419178082	1	0.32	266	86.36
33.753424658	1	0.32	267	86.69
33.915068493	2	0.65	269	87.34
34	39	12.66	308	100.00

Vacation	Frequency	Percent	Cumulative Frequency	Cumulative Percent
15	120	38.96	120	38.96
20	188	61.04	308	100.00



ıry	Status	City	State	Country	BirthDate	HireDate	TermDate	YearsEmp	Bonus	Vacation
75	M	Sydney		AU	-1805	9132	.	34	3299.0625	20
00	M	Sydney		AU	-3296	9132	.	34	1011	20
50	S	Melbourne		AU	-3523	9132	.	34	1007.625	20
25	M	Sydney		AU	-3614	9132	.	34	1173.1875	20
25	M	Sydney		AU	-3528	9132	.	34	1036.6875	20
00	M	Sydney		AU	-3439	9132	.	34	997.5	20
00	M	Sydney		AU	-3328	9132	.	34	994.5	20
50	M	Sydney		AU	-3452	9132	.	34	1143.375	20
25	M	Sydney		AU	-3562	9132	.	34	1062.9375	20
75	O	Miami-Dade	FL	US	84	9132	.	34	1816.3125	20
50	S	Philadelphia	PA	US	-3579	9132	.	34	1289.625	20
50	M	Philadelphia	PA	US	-1630	9132	.	34	6048.375	20
50	S	Philadelphia	PA	US	-3449	9132	.	34	1363.875	20
75	S	Melbourne		AU	-3475	9132	.	34	1347.5625	20
50	M	San Diego	CA	US	-3526	9132	.	34	1000.875	20

/* Analyzing Data */

/* Orders Frequency Analysis */

```
proc freq data=profit_country;
        tables Order_Date;
        format Order_Date monname.;
```

```
run;


ods noproctitle;
title "Number of Days by Month";
title2 "And Customer Continent / Order Source";
proc freq data=profit_country order=freq;
        tables Order_Date / nocum;
        format Order_Date monname.;
        tables Customer_Continent*Order_Source / norow nocol;
run;
```

The FREQ Procedure

Order_Date	Frequency	Percent	Cumulative Frequency	Cumulative Percent
January	750	6.96	750	6.96
February	700	6.49	1450	13.45
March	728	6.75	2178	20.20
April	886	8.22	3064	28.42
May	964	8.94	4028	37.36
June	913	8.47	4941	45.83
July	981	9.10	5922	54.93
August	1026	9.52	6948	64.45
September	741	6.87	7689	71.32
October	908	8.42	8597	79.74
November	984	9.13	9581	88.87
December	1200	11.13	10781	100.00

Number of Days by Month
And Customer Continent / Order Source

Order_Date	Frequency	Percent
December	1200	11.13
August	1026	9.52
November	984	9.13
July	981	9.10
May	964	8.94
June	913	8.47
October	908	8.42
April	886	8.22
January	750	6.96
September	741	6.87
March	728	6.75
February	700	6.49

Table of Customer_Continent by Order_Source					
	Order_Source				
Customer_Continent	Retail	Internet	Phone	Unknown	Total
Europe	6002 55.67	848 7.87	741 6.87	3 0.03	7594 70.44
North America	1772 16.44	384 3.56	273 2.53	0 0.00	2429 22.53
Australia/Pacific	559 5.19	106 0.98	67 0.62	0 0.00	732 6.79
Africa	0 0.00	9 0.08	5 0.05	0 0.00	14 0.13
Asia	0 0.00	9 0.08	3 0.03	0 0.00	12 0.11
Total	8333 77.29	1356 12.58	1089 10.10	3 0.03	10781 100.00

```
title "Days Ship to by Country";
title2 "Phone Order";
proc means data=profit_country min max mean maxdec=0;
        var Ship_Days;
        class Country_Name;
        where Ship_Days>0 and Order_Source='Phone';
run;
```

Days Ship to by Country
Phone Order

Analysis Variable : Ship_Days				
Country_Name	N Obs	Minimum	Maximum	Mean
Australia	67	1	10	4
Austria	1	4	4	4
Belgium	36	1	5	3
Canada	16	1	4	3
Croatia	1	3	3	3
Denmark	26	1	7	3
Finland	9	1	5	3
France	112	1	7	3
Germany	123	1	9	3
Greece	2	3	4	4
Hungary	1	2	2	2
Ireland	1	2	2	2
Italy	128	1	12	3
Luxembourg	2	1	2	2
Netherlands	82	1	5	3
Norway	4	1	7	4
Portugal	5	2	5	3
South Africa	5	1	4	2
Spain	113	1	11	3
Sweden	7	1	4	3
Switzerland	7	2	5	3
Turkey	3	3	5	4
United Kingdom	80	1	19	4
United States	257	1	9	3

```
%let os=Retail;
title "Days Ship to by Country";
title2 "&os Order";
proc means data=profit_country min max mean maxdec=0;
        var Ship_Days;
        class Country_Name;
        where Ship_Days>0 and Order_Source="&os";
run;
```

Days Ship to by Country
Retail Order

Analysis Variable : Ship_Days				
Country_Name	N Obs	Minimum	Maximum	Mean
Australia	35	1	17	5
Belgium	18	1	15	5
Denmark	6	1	13	6
France	52	1	16	5
Germany	74	1	15	5
Italy	67	1	17	5
Netherlands	30	1	14	4
Spain	48	1	14	5
United Kingdom	54	1	18	5
United States	124	1	22	5

```
title "Days Ship to by Country";
proc means data=profit_country min max mean maxdec=0;
        var Ship_Days;
        class Order_Source Country_Name;
        where Ship_Days>0;
run;
```

Days Ship to by Country

Analysis Variable : Ship_Days					
Order_Source	Country_Name	N Obs	Minimum	Maximum	Mean
Internet	Australia	106	1	6	3
	Austria	11	1	13	5
	Belgium	32	1	6	3
	Canada	28	1	6	4
	Czech Republic	1	1	1	1
	Denmark	19	1	6	3
	Finland	3	1	8	4
	France	165	1	15	4
	Germany	128	1	10	4
	Ireland	1	3	3	3
	Israel	1	4	4	4
	Italy	142	1	6	4
	Netherlands	83	1	12	4
	Norway	4	3	6	5
	Portugal	7	2	6	5
	Russia	1	4	4	4
	Slovenia	2	2	3	3
	South Africa	9	2	6	4
	Spain	141	1	20	4
	Sweden	2	3	4	4
	Switzerland	6	3	6	5
	Turkey	5	1	5	2
	United Arab Emirates	2	1	1	1
	United Kingdom	101	1	12	4
	United States	356	1	9	3
Phone	Australia	67	1	10	4
	Austria	1	4	4	4
	Belgium	36	1	5	3
	Canada	16	1	4	3
	Croatia	1	3	3	3
	Denmark	26	1	7	3
	Finland	9	1	5	3
	France	112	1	7	3
	Germany	123	1	9	3

```
proc sort data=profit_country out=profit_sort;
        by Order_Source;
run;
```

```
title "Days Ship to by Country";
proc means data=profit_sort min max mean maxdec=0;
        var Ship_Days;
        class Country_Name;
        where Ship_Days>0;
        by Order_Source;
run;
```

Days Ship to by Country

Order_Source=Internet

Analysis Variable : Ship_Days				
Country_Name	N Obs	Minimum	Maximum	Mean
Australia	106	1	6	3
Austria	11	1	13	5
Belgium	32	1	6	3
Canada	28	1	6	4
Czech Republic	1	1	1	1
Denmark	19	1	6	3
Finland	3	1	8	4
France	165	1	15	4
Germany	128	1	10	4
Ireland	1	3	3	3
Israel	1	4	4	4
Italy	142	1	6	4
Netherlands	83	1	12	4
Norway	4	3	6	5
Portugal	7	2	6	5
Russia	1	4	4	4
Slovenia	2	2	3	3
South Africa	9	2	6	4
Spain	141	1	20	4
Sweden	2	3	4	4
Switzerland	6	3	6	5
Turkey	5	1	5	2
United Arab Emirates	2	1	1	1
United Kingdom	101	1	12	4
United States	356	1	9	3

Order_Source=Phone

Analysis Variable : Ship_Days				
Country_Name	N Obs	Minimum	Maximum	Mean
Australia	67	1	10	4
Austria	1	4	4	4
Belgium	36	1	5	3
Canada	16	1	4	3
Croatia	1	3	3	3
Denmark	26	1	7	3
Finland	9	1	5	3
France	112	1	7	3
Germany	123	1	9	3
Greece	2	3	4	4
Hungary	1	2	2	2
Ireland	1	2	2	2
Italy	128	1	12	3
Luxembourg	2	1	2	2
Netherlands	82	1	5	3
Norway	4	1	7	4
Portugal	5	2	5	3
South Africa	5	1	4	2
Spain	113	1	11	3
Sweden	7	1	4	3
Switzerland	7	2	5	3
Turkey	3	3	5	4
United Kingdom	80	1	19	4
United States	257	1	9	3

Order_Source=Retail

Analysis Variable : Ship_Days				
Country_Name	N Obs	Minimum	Maximum	Mean
Australia	35	1	17	5
Belgium	18	1	15	5
Denmark	6	1	13	6
France	52	1	16	5
Germany	74	1	15	5
Italy	67	1	17	5
Netherlands	30	1	14	4
Spain	48	1	14	5
United Kingdom	54	1	18	5
United States	124	1	22	5

Order_Source=Unknown

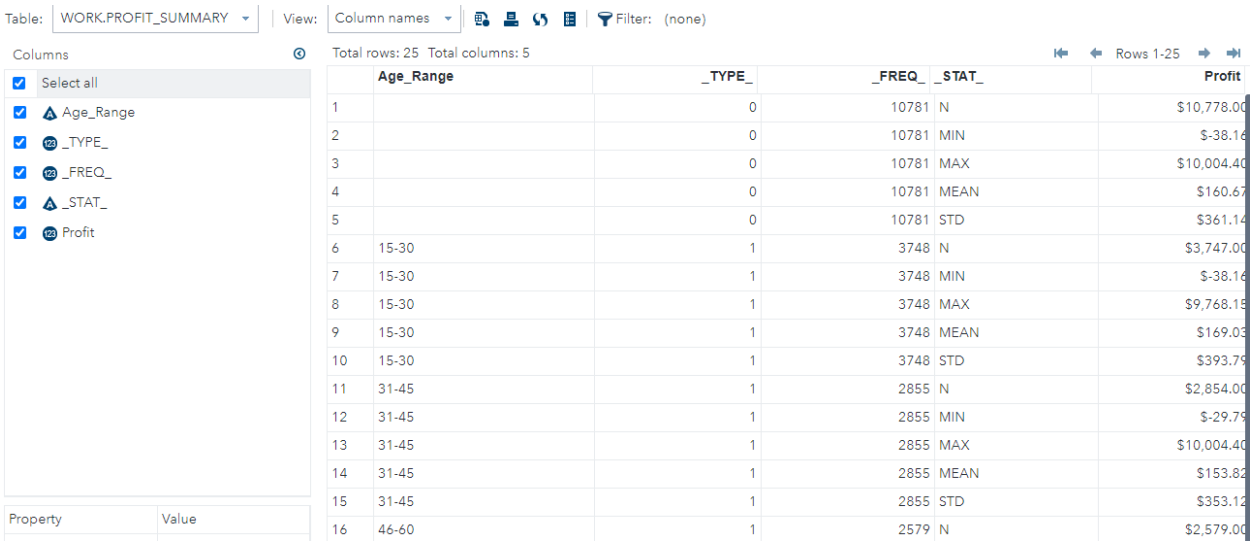
Analysis Variable : Ship_Days				
Country_Name	N Obs	Minimum	Maximum	Mean
Germany	1	5	5	5

/* Profit Analysis by Customer Age */

proc means data=profit_country noprint;

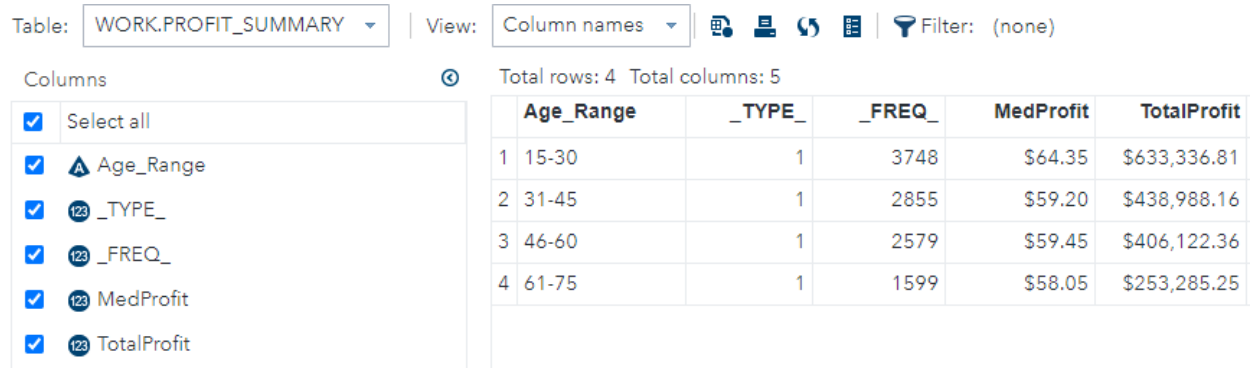
var Profit;

```
        class Age_Range;
        output out=profit_summary;
run;
```



	Age_Range	_TYPE_	_FREQ_	_STAT_	Profit
1		0	10781	N	$10,778.00
2		0	10781	MIN	$-38.16
3		0	10781	MAX	$10,004.40
4		0	10781	MEAN	$160.67
5		0	10781	STD	$361.14
6	15-30	1	3748	N	$3,747.00
7	15-30	1	3748	MIN	$-38.16
8	15-30	1	3748	MAX	$9,768.18
9	15-30	1	3748	MEAN	$169.03
10	15-30	1	3748	STD	$393.79
11	31-45	1	2855	N	$2,854.00
12	31-45	1	2855	MIN	$-29.79
13	31-45	1	2855	MAX	$10,004.40
14	31-45	1	2855	MEAN	$153.82
15	31-45	1	2855	STD	$353.12
16	46-60	1	2579	N	$2,579.00

```
proc means data=profit_country noprint;
        var Profit;
        class Age_Range;
        output out=profit_summary Median=MedProfit Sum=TotalProfit;
        ways 1;
run;
```



	Age_Range	_TYPE_	_FREQ_	MedProfit	TotalProfit
1	15-30	1	3748	$64.35	$633,336.81
2	31-45	1	2855	$59.20	$438,988.16
3	46-60	1	2579	$59.45	$406,122.36
4	61-75	1	1599	$58.05	$253,285.25

```
title "Profit by Customer Age Range";
proc print data=profit_summary noobs label;
        var Age_Range TotalProfit MedProfit;
        label Age_Range="Age Range"
                TotalProfit="Total Profit"
```

```
                MedProfit="Median Profit per Order";
        format TotalProfit MedProfit dollar10.;
run;
```

Profit by Customer Age Range

Age Range	Total Profit	Median Profit per Order
15-30	$633,337	$64
31-45	$438,988	$59
46-60	$406,122	$59
61-75	$253,285	$58

```
/*Programming Question 2.04
If necessary, start SAS Studio. Open p105q1.sas from the programs folder.
Identify and fix the errors so that the program creates the output shown below.
Run the program.
What is the label for nHome in the sashelp.baseball table?
Which option must be added so that the descriptive column heading text appears in the report?
How many total home runs did Baltimore have?
Which statement must be added so that the player's name replaces the default OBS column?
*/

proc sort data=sashelp.baseball out=baseball_sort;
        by Team;
run;

title "Baseball Team Statistics";
proc print data=baseball_sort label;
        by Team;
        var Name Position YrMajor nAtBat nHits nHome;
        sum nAtBat nHits nHome;
        label name="Player's Name"
                position="Position(s) in 1986"
```

```
                    YrMajor="Years in the Major League"
                    nAtBat="Times at Bat in 1986"
                    nHits="Hits in 1986"
                    nHome="Home Runs in 1986";
run;

*Solution;
proc sort data=sashelp.baseball out=baseball_sort;
   by Team Name;
run;

title "Baseball Team Statistics";
proc print data=baseball_sort label;
   by Team;
   id Name;
   var Position YrMajor nAtBat nHits nHome;
   sum nAtBat nHits nHome;
run;
```

Baseball Team Statistics

Team at the End of 1986=Atlanta

Player's Name	Position(s) in 1986	Years in the Major Leagues	Times at Bat in 1986	Hits in 1986	Home Runs in 1986
Harper, Terry	OF	7	265	68	8
Horner, Bob	1B	9	517	141	27
Hubbard, Glenn	2B	9	408	94	4
Moreno, Omar	RF	12	359	84	4
Murphy, Dale	CF	11	614	163	29
Oberkfell, Ken	3B	10	503	136	5
Ramirez, Rafael	S3	7	496	119	8
Sample, Billy	OF	9	200	57	6
Simmons, Ted	UT	19	127	32	4
Thomas, Andres	SS	2	323	81	6
Virgil, Ozzie	C	7	359	80	15
Team			4171	1055	116

Team at the End of 1986=Baltimore

Player's Name	Position(s) in 1986	Years in the Major Leagues	Times at Bat in 1986	Hits in 1986	Home Runs in 1986
Beniquez, Juan	UT	15	343	103	6
Bonilla, Juan	2B	5	284	69	1
Dempsey, Rick	C	18	327	68	13
Dwyer, Jim	DO	14	160	39	8
Lacy, Lee	RF	15	491	141	11
Lynn, Fred	CF	13	397	114	23
Murray, Eddie	1B	10	495	151	17
O'Malley, Tom	3B	5	181	46	1
Rayford, Floyd	3B	6	210	37	8
Ripken, Cal	SS	6	627	177	25
Sheets, Larry	DH	3	338	92	18
Shelby, John	OF	6	404	92	11
Traber, Jim	UT	2	212	54	13
Wiggins, Alan	2B	6	239	60	0
Young, Mike	LF	5	369	93	9
Team			5077	1336	164

/* Programming Question 2.05

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

Write and submit a program to analyze the number of employees in the cr.employee table by City, Department, and JobTitle.

Which city has the highest number of employees?

What percentage of all employees are in the Sales Department? Note: Type your answer exactly the way the value is displayed

```
How many unique values of JobTitle are in the employee table?
*/

proc freq data=cr.employee;
        tables City Department JobTitle;
run;

proc sql;
        select count(distinct JobTitle)
        from cr.employee;
quit;

*Solution;
proc freq data=cr.employee order=freq nlevels;
   tables City Department JobTitle / nocum;
run;
```

The FREQ Procedure

Number of Variable Levels		
Variable	Label	Levels
City	City	5
Department		17
JobTitle		132

City		
City	Frequency	Percent
San Diego	112	26.42
Miami-Dade	109	25.71
Philadelphia	95	22.41
Sydney	67	15.80
Melbourne	41	9.67

Department	Frequency	Percent
Sales	201	47.41
Administration	34	8.02
Stock & Shipping	26	6.13
IS	25	5.90
Marketing	20	4.72
Group HR Management	18	4.25
Purchasing	18	4.25
Accounts	17	4.01
Logistics Management	14	3.30
Concession Management	11	2.59
Sales Management	11	2.59
Accounts Management	9	2.12
Engineering	9	2.12
Executives	4	0.94
Group Financials	3	0.71
Secretary of the Board	2	0.47
Strategy	2	0.47

JobTitle	Frequency	Percent
Sales Rep. I	63	14.86
Sales Rep. II	46	10.85

/*Programming Question 2.06

If necessary, start SAS Studio and submit libname.sas. Write and submit a program to do the following:

Analyze frequency counts for the cr.profit table.

Create a two-way frequency table that includes the frequency count and percent for each Order_Date and Order_Source.

Display Order_Date using the MONNAME format so there is one row per month in the table.

How many retail orders were in December?

What percentage of all orders were retail? Note: Type your answer exactly as the value appears in the report.

*/

```
proc freq data=cr.profit;
        tables Order_Date*Order_Source / norow nocol nocum;
        format Order_date monname.;
run;

*Solution;
proc freq data=cr.profit;
   format Order_Date monname.;
   tables Order_Date*Order_Source / nocol norow ;
run;
```

The FREQ Procedure

Frequency Percent	Table of Order_Date by Order_Source				
	Order_Source				
Order_Date	Internet	Phone	Retail	Unknown	Total
January	95 0.88	99 0.92	556 5.16	0 0.00	750 6.96
February	85 0.79	62 0.58	553 5.13	0 0.00	700 6.49
March	81 0.75	74 0.69	573 5.31	0 0.00	728 6.75
April	128 1.19	110 1.02	648 6.01	0 0.00	886 8.22
May	137 1.27	109 1.01	718 6.66	0 0.00	964 8.94
June	149 1.38	99 0.92	665 6.17	0 0.00	913 8.47
July	143 1.33	124 1.15	714 6.62	0 0.00	981 9.10
August	140 1.30	120 1.11	766 7.11	0 0.00	1026 9.52
September	54 0.50	48 0.45	639 5.93	0 0.00	741 6.87
October	102 0.95	64 0.59	742 6.88	0 0.00	908 8.42
November	102 0.95	86 0.80	796 7.38	0 0.00	984 9.13
December	140 1.30	94 0.87	963 8.93	3 0.03	1200 11.13
Total	1356 12.58	1089 10.10	8333 77.29	3 0.03	10781 100.00

/*Programming Question 2.07

If necessary, start SAS Studio and submit libname.sas. Write and submit a program to do the following:

Calculate summary statistics for the cr.employee table.

Subset the rows to include only the Sales Department.

Calculate the sum, mean, minimum, and maximum of Salary for each value of JobTitle. Round values to the nearest whole number.

What is the total salary for all Sales employees combined? Note: Type the value exactly as it appears.

What is the mean salary for Sales Rep. IV employees? Note: Type your answer exactly as shown.

*/

```
proc means data=cr.employee sum mean min max maxdec=0;
        where Department="Sales";
```

var Salary;

class JobTitle;

ways 0 1;

run;

The MEANS Procedure

Analysis Variable : Salary				
N Obs	Sum	Mean	Minimum	Maximum
201	6910144	34379	28388	45756

Analysis Variable : Salary					
JobTitle	N Obs	Sum	Mean	Minimum	Maximum
Sales Rep. I	63	2086744	33123	28388	40294
Sales Rep. II	46	1577631	34296	32575	44988
Sales Rep. III	34	1251469	36808	35031	45756
Sales Rep. IV	16	633088	39568	37881	41231
Temp. Sales Rep.	24	787975	32832	31275	34350
Trainee	18	573238	31847	30019	38681

/*Programming Question 2.08

If necessary, start SAS Studio and submit libname.sas. Write and submit a program to do the following:

Use PROC MEANS to read cr.employee and create a summary table named salary_summary. Do not create a report.

Compute statistics for Salary based on the values of Department and City.

In the output table, include a column named TotalSalary that is the sum of Salary and a column named AvgSalary that is the mean of Salary.

In the output table, include one row for each unique combination of Department and City.

How many rows are in the salary_summary table?

What is the value of AvgSalary for the Administration Department in Sydney?

*/

proc means data=cr.employee noprint;

var Salary;

class Department City;

output out=salary_summary Mean=AvgSalary Sum=TotalSalary;

ways 2;

```
run;


title "Salary by Department and City";
proc print data=salary_summary noobs label;
        var Deparment City TotalSalary AvgSalary;
        label Department="Department"
                City="City"
                TotalSalary="Total Salary"
                AvgSalary="Mean Salary per Employee";
        format TotalSalary AvgSalary dollar10.;
run;


*Solution;
proc means data=cr.employee noprint;
   var Salary;
   class Department City;
   output out=salary_summary mean=AvgSalary sum=TotalSalary;
   ways 2;
run;
```

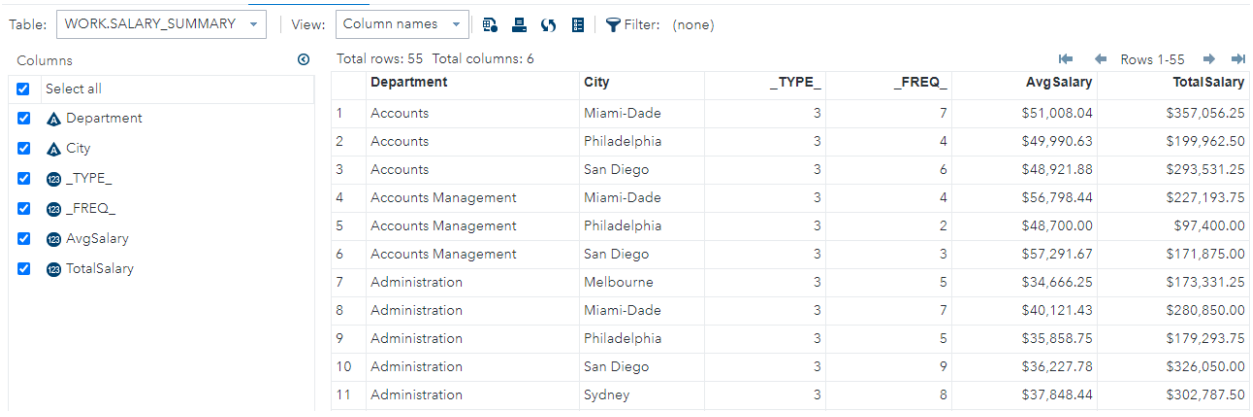


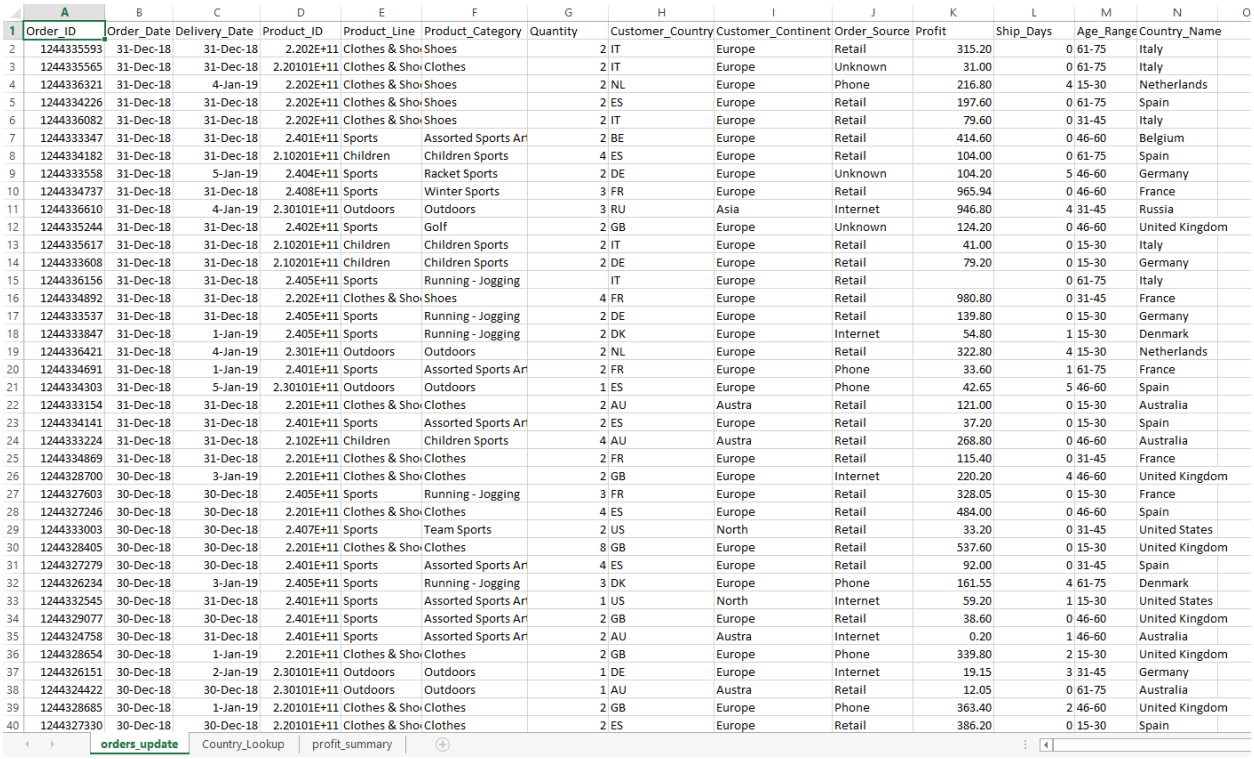
Table: WORK.SALARY_SUMMARY | View: Column names | Filter: (none)

Columns: Select all, Department, City, _TYPE_, _FREQ_, AvgSalary, TotalSalary

Total rows: 55 Total columns: 6 — Rows 1-55

	Department	City	_TYPE_	_FREQ_	Avg Salary	Total Salary
1	Accounts	Miami-Dade	3	7	$51,008.04	$357,056.25
2	Accounts	Philadelphia	3	4	$49,990.63	$199,962.50
3	Accounts	San Diego	3	6	$48,921.88	$293,531.25
4	Accounts Management	Miami-Dade	3	4	$56,798.44	$227,193.75
5	Accounts Management	Philadelphia	3	2	$48,700.00	$97,400.00
6	Accounts Management	San Diego	3	3	$57,291.67	$171,875.00
7	Administration	Melbourne	3	5	$34,666.25	$173,331.25
8	Administration	Miami-Dade	3	7	$40,121.43	$280,850.00
9	Administration	Philadelphia	3	5	$35,858.75	$179,293.75
10	Administration	San Diego	3	9	$36,227.78	$326,050.00
11	Administration	Sydney	3	8	$37,848.44	$302,787.50

```
/* Exporting Data */
proc export data=profit_country outfile="&path/output/orders_update.csv" dbms=csv replace;
run;

proc export data=profit_country outfile="&path/output/orders_update.xlsx" dbms=xlsx replace;
run;

libname outxl xlsx "&path/output/orders_update.xlsx";

data outxl.Orders_Update;
	set Profit_Country;
run;

data outxl.Country_Lookup;
	set Country_Clean;
run;

proc means data=profit noprint;
	var profit;
	class Age_Range;
	ways 1;
	output out=outxl.profit_summary;
run;

libname outxl clear;
```



Order_ID	Order_Date	Delivery_Date	Product_ID	Product_Line	Product_Category	Quantity	Customer_Country	Customer_Continent	Order_Source	Profit	Ship_Days	Age_Range	Country_Name
1244335593	31-Dec-18	31-Dec-18	2.202E+11	Clothes & Sho	Shoes	2	IT	Europe	Retail	315.20	0	61-75	Italy
1244335565	31-Dec-18	31-Dec-18	2.20101E+11	Clothes & Sho	Clothes	2	IT	Europe	Unknown	31.00	0	61-75	Italy
1244336321	31-Dec-18	4-Jan-19	2.202E+11	Clothes & Sho	Shoes	2	NL	Europe	Phone	216.80	4	15-30	Netherlands
1244334226	31-Dec-18	31-Dec-18	2.202E+11	Clothes & Sho	Shoes	2	ES	Europe	Retail	197.60	0	61-75	Spain
1244336082	31-Dec-18	31-Dec-18	2.202E+11	Clothes & Sho	Shoes	2	IT	Europe	Retail	79.60	0	31-45	Italy
1244333347	31-Dec-18	31-Dec-18	2.401E+11	Sports	Assorted Sports Ar	2	BE	Europe	Retail	414.60	0	46-60	Belgium
1244334182	31-Dec-18	31-Dec-18	2.10201E+11	Children	Children Sports	4	ES	Europe	Retail	104.00	0	61-75	Spain
1244333558	31-Dec-18	5-Jan-19	2.404E+11	Sports	Racket Sports	2	DE	Europe	Unknown	104.20	5	46-60	Germany
1244334737	31-Dec-18	31-Dec-18	2.408E+11	Sports	Winter Sports	3	FR	Europe	Retail	965.94	0	46-60	France
1244336610	31-Dec-18	4-Jan-19	2.30101E+11	Outdoors	Outdoors	3	RU	Asia	Internet	946.80	4	31-45	Russia
1244335244	31-Dec-18	31-Dec-18	2.402E+11	Sports	Golf	2	GB	Europe	Unknown	124.20	0	46-60	United Kingdom
1244335617	31-Dec-18	31-Dec-18	2.10201E+11	Children	Children Sports	2	IT	Europe	Retail	41.00	0	15-30	Italy
1244333608	31-Dec-18	31-Dec-18	2.10201E+11	Children	Children Sports	2	DE	Europe	Retail	79.20	0	15-30	Germany
1244336156	31-Dec-18	31-Dec-18	2.405E+11	Sports	Running - Jogging		IT	Europe	Retail		0	61-75	Italy
1244334892	31-Dec-18	31-Dec-18	2.202E+11	Clothes & Sho	Shoes	4	FR	Europe	Retail	980.80	0	31-45	France
1244333537	31-Dec-18	31-Dec-18	2.405E+11	Sports	Running - Jogging	2	DE	Europe	Retail	139.80	0	15-30	Germany
1244333847	31-Dec-18	1-Jan-19	2.405E+11	Sports	Running - Jogging	2	DK	Europe	Internet	54.80	1	15-30	Denmark
1244336421	31-Dec-18	4-Jan-19	2.301E+11	Outdoors	Outdoors	2	NL	Europe	Retail	322.80	4	15-30	Netherlands
1244334691	31-Dec-18	1-Jan-19	2.401E+11	Sports	Assorted Sports Ar	2	FR	Europe	Phone	33.60	1	61-75	France
1244334303	31-Dec-18	5-Jan-19	2.30101E+11	Outdoors	Outdoors	1	ES	Europe	Phone	42.65	5	46-60	Spain
1244333154	31-Dec-18	31-Dec-18	2.201E+11	Clothes & Sho	Clothes	2	AU	Austra	Retail	121.00	0	15-30	Australia
1244334141	31-Dec-18	31-Dec-18	2.401E+11	Sports	Assorted Sports Ar	2	ES	Europe	Retail	37.20	0	15-30	Spain
1244333224	31-Dec-18	31-Dec-18	2.102E+11	Children	Children Sports	4	AU	Austra	Retail	268.80	0	46-60	Australia
1244334869	31-Dec-18	31-Dec-18	2.201E+11	Clothes & Sho	Clothes	2	FR	Europe	Retail	115.40	0	31-45	France
1244328700	30-Dec-18	3-Jan-19	2.201E+11	Clothes & Sho	Clothes	2	GB	Europe	Internet	220.20	4	46-60	United Kingdom
1244327603	30-Dec-18	30-Dec-18	2.405E+11	Sports	Running - Jogging	3	FR	Europe	Retail	328.05	0	15-30	France
1244327246	30-Dec-18	30-Dec-18	2.201E+11	Clothes & Sho	Clothes	4	ES	Europe	Retail	484.00	0	46-60	Spain
1244333003	30-Dec-18	30-Dec-18	2.407E+11	Sports	Team Sports	2	US	North	Retail	33.20	0	31-45	United States
1244328405	30-Dec-18	30-Dec-18	2.201E+11	Clothes & Sho	Clothes	8	GB	Europe	Retail	537.60	0	15-30	United Kingdom
1244327279	30-Dec-18	30-Dec-18	2.401E+11	Sports	Assorted Sports Ar	4	ES	Europe	Retail	92.00	0	31-45	Spain
1244326234	30-Dec-18	3-Jan-19	2.405E+11	Sports	Running - Jogging	3	DK	Europe	Phone	161.55	4	61-75	Denmark
1244332545	30-Dec-18	31-Dec-18	2.401E+11	Sports	Assorted Sports Ar	1	US	North	Internet	59.20	1	15-30	United States
1244329077	30-Dec-18	30-Dec-18	2.401E+11	Sports	Assorted Sports Ar	2	GB	Europe	Retail	38.60	0	46-60	United Kingdom
1244324758	30-Dec-18	31-Dec-18	2.401E+11	Sports	Assorted Sports Ar	2	AU	Austra	Internet	0.20	1	46-60	Australia
1244328654	30-Dec-18	1-Jan-19	2.201E+11	Clothes & Sho	Clothes	2	GB	Europe	Phone	339.80	2	15-30	United Kingdom
1244326151	30-Dec-18	2-Jan-19	2.30101E+11	Outdoors	Outdoors	1	DE	Europe	Internet	19.15	3	31-45	Germany
1244324422	30-Dec-18	30-Dec-18	2.30101E+11	Outdoors	Outdoors	1	AU	Austra	Retail	12.05	0	61-75	Australia
1244328685	30-Dec-18	1-Jan-19	2.20101E+11	Clothes & Sho	Clothes	2	GB	Europe	Phone	363.40	2	46-60	United Kingdom
1244327330	30-Dec-18	30-Dec-18	2.20101E+11	Clothes & Sho	Clothes	2	ES	Europe	Retail	386.20	0	15-30	Spain

orders_update | Country_Lookup | profit_summary

/* Exporting PDF Results from Orders Exploration */

ods pdf file="&path/output/orders_validation.pdf";

title "Orders with Order Date after Delivery Date";

proc print data=cr.orders;

where Order_Date>Delivery_Date;

var Order_ID Order_Date Delivery_Date;

run;

title "Examine values of Numeric Columns in Orders";

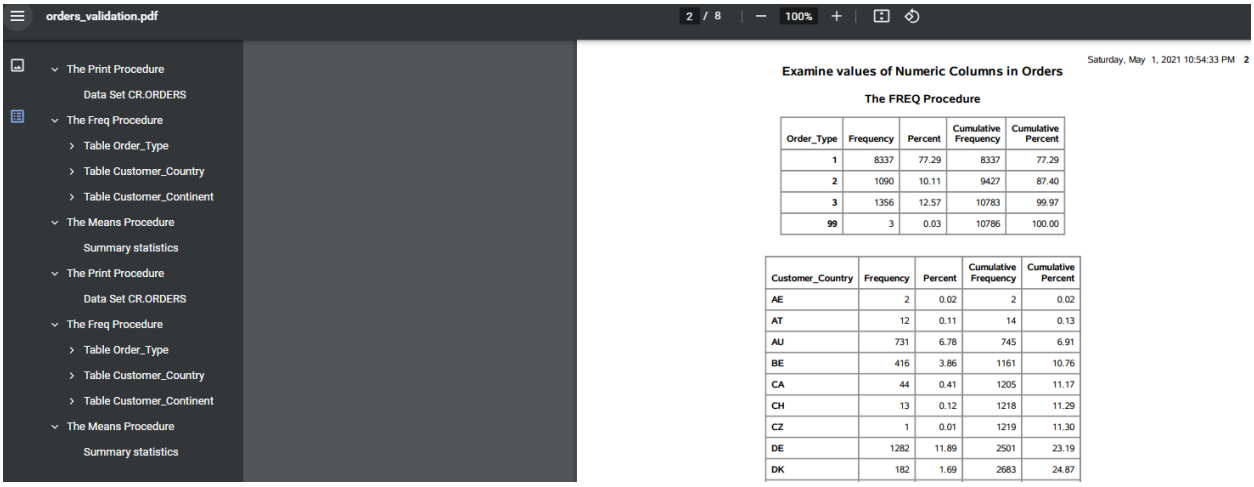
proc freq data=cr.orders;

tables Order_Type Customer_Country Customer_Continent;

run;

title "Examine values of Categorical Columns in Orders";

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

ods pdf close;
```

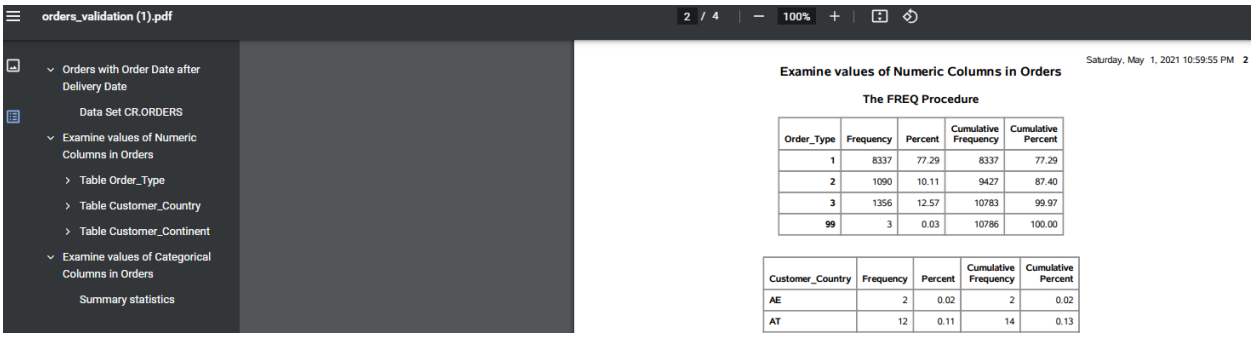


Order_Type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	8337	77.29	8337	77.29
2	1090	10.11	9427	87.40
3	1356	12.57	10783	99.97
99	3	0.03	10786	100.00

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

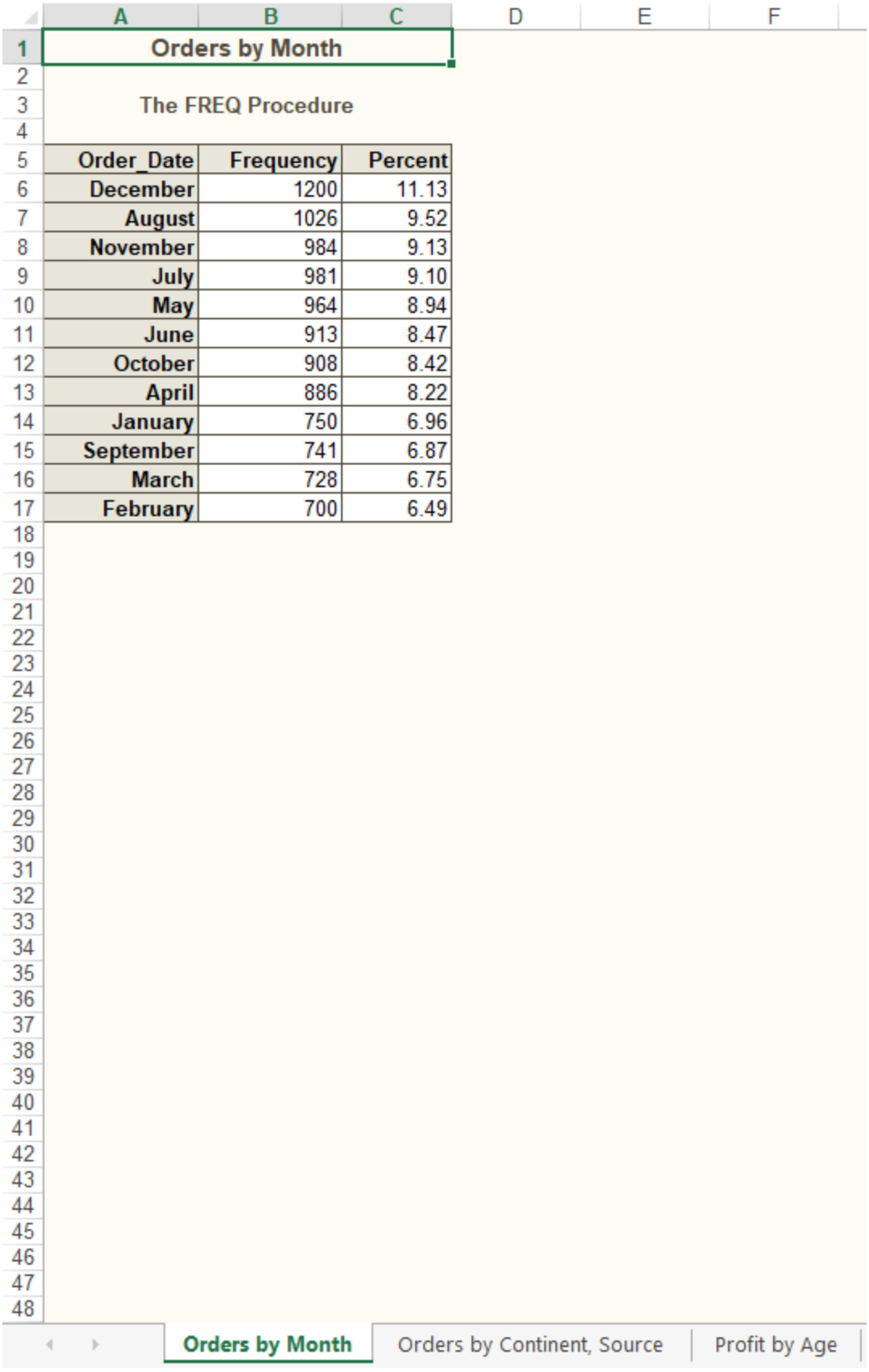
```
/* Exporting PDF Results from Orders Exploration with customized label */

ods pdf file="&path/output/orders_validation.pdf" pdftoc=1;

ods proclabel "Orders with Order Date after Delivery Date";
title "Orders with Order Date after Delivery Date";
proc print data=cr.orders;
        where Order_Date>Delivery_Date;
        var Order_ID Order_Date Delivery_Date;
run;

ods proclabel "Examine values of Numeric Columns in Orders";
title "Examine values of Numeric Columns in Orders";
proc freq data=cr.orders;
        tables Order_Type Customer_Country Customer_Continent;
run;
```

```
ods proclabel "Examine values of Categorical Columns in Orders";
title "Examine values of Categorical Columns in Orders";
proc means data=cr.orders maxdec=0;
        var Quantity Retail_Price Cost_Price;
run;

ods pdf close;
```




```
/* Exporting Excel Results from Orders Exploration with customized label */

ods excel file="&path/output/Orders_Analysis.xlsx"
        options(embedded_titles="yes") style=analysis;

ods excel options(sheet_name="Orders by Month");
title "Orders by Month";
proc freq data=profit_country order=freq;
        tables Order_Date / nocum;
        format Order_Date monname.;
run;

ods excel options(sheet_name="Orders by Continent, Source");
title "Orders by Customer Continent/Order Source";
proc freq data=profit_country order=freq;
```

```
        tables Customer_Continent*Order_Source / norow nocol;
run;

proc means data=profit noprint;
        var Profit;
        class Age_Range;
        output out=profit_summary;
run;

proc means data=profit Median sum noprint;
        var profit;
        class Age_Range;
        output out=profit_summary Median=MedProfit sum=TotalProfit;
        ways 1;
run;

ods excel options(sheet_name="Profit by Age");
title "Profit by Customer Age Range";
proc print data=profit_summary label noobs;
        var Age_Range TotalProfit MedProfit;
        label Age_Range="Age Range"
                TotalProfit="Total Profit"
                MedProfit="Median Profit per Order";
        format TotalProfit MedProfit dollar10.;
run;

ods excel close;
```

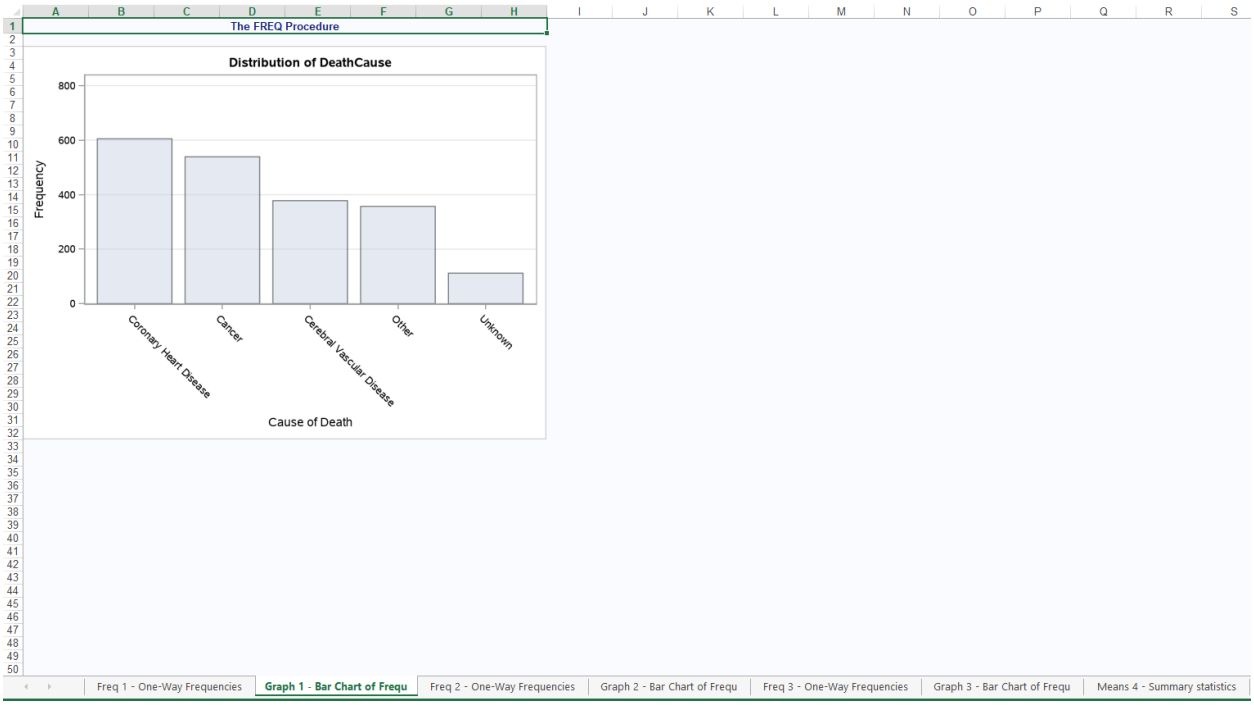


Orders by Month

The FREQ Procedure

Order_Date	Frequency	Percent
December	1200	11.13
August	1026	9.52
November	984	9.13
July	981	9.10
May	964	8.94
June	913	8.47
October	908	8.42
April	886	8.22
January	750	6.96
September	741	6.87
March	728	6.75
February	700	6.49

```
/*Programming Question 2.09
If necessary, start SAS Studio. Open p106q1.sas from the programs folder.
Identify and fix the errors.
Run the code.
In the navigation pane, select the heart.xlsx file in the output folder.
Click the download button and open the Excel file.
Are the titles specified in the TITLE statements included in the spreadsheets?
How many worksheets are included in the Excel file?
*/

ods graphics on;
ods noproctitle;

ods excel file="&path/heart.xlsx";
title "Distribution of Patient Status";
proc freq data=sashelp.heart order=freq;
        tables DeathCause Chol_Status BP_Status / nocum plots=freqplot;
run;

title "Summary of Measures for Patients";
proc means data=sashelp.heart mean;
        var AgeAtDeath Cholesterol Weight Smoking;
        class Sex;
run;
ods excel close;
ods proctitle;
```



/*Programming Question 2.10

If necessary, start SAS Studio. Open p106q2.sas from the programs folder. Modify the program to do the following:

Create a PDF file named truck.pdf in the output folder that combines the results of the two procedures

without a page break between the reports.

Apply the Journal style.

In the navigation pane, find the truck.pdf file in the output folder.

Right-click the file and select Download File.

What is the label for the first bookmark in the PDF table of contents?

PROC FREQ Table

The Freq Procedure

PROC FREQ Output

Do both reports fit on a single page?

yes

no

*/

```
ods pdf file="&path/output/truck.pdf" pdftoc=1;
ods noproctitle;
title "Truck Summary";
title2 "SASHELP.CARS Table";

proc freq data=sashelp.cars;
        where Type="Truck";
        tables Make / nocum;
run;

proc print data=sashelp.cars;
        where Type="Truck";
        id Make;
        var Model MSRP MPG_City MPG_Highway;
run;

ods pdf close;

*Solution;
ods pdf file="&path/truck.pdf" style=journal startpage=no;
ods noproctitle;
title "Truck Summary";
title2 "SASHELP.CARS Table";

proc freq data=sashelp.cars;
         where Type="Truck";
         tables Make / nocum;
run;
```

```
proc print data=sashelp.cars;
        where Type="Truck";
        id Make;
        var Model MSRP MPG_City MPG_Highway;
run;
ods pdf close;


/*Programming Question 2.11
If necessary, start SAS Studio and submit libname.sas.
Write a new program that uses PROC EXPORT to export cr.employee_current to the output folder as a comma-delimited file.
In the navigation pane, find the exported text file in the output folder.
Right-click the file and select Properties.
What is the file size, rounded to the nearest KB? Note: Type a number for your answer.
In the navigation pane, right-click the file and select View File as Text. Which fields are enclosed in quotation marks?
Department
Salary
Name
JobTitle
*/


proc export data=cr.employee_current outfile="&path/output/employee_current.csv" dbms=csv
replace;
run;
```