

BUAN 6337 Homework 1_Group 9

Question 1

A gourmet pizza restaurant is considering adding new toppings to its menu. Each month they survey 10 customers about their preferences for three different toppings. The restaurant wants data on several different toppings. So they don't always ask about the same three toppings. Customers rate each topping on a scale of 1(would never order) to 5 (would order often). The restaurant wants to compute average ratings for all toppings, so the ratings variables need to be numeric. The raw data file Pizza.csv has variables for the respondent's ID, and the ratings for five different toppings: arugula, pine nuts, roasted butternut squash, shrimp, and grilled eggplant. The first two digits in the ID correspond to the month of the survey.

- Examine the raw data file Pizza.csv and read it into SAS using the IMPORT procedure. Print the data set (on the results screen). Print a report that describes the contents of the data set to make sure all the variables are the correct type.
- Open the raw data file in a simple editor like WordPad and compare the data values to the output from part b) to make sure that they were read correctly into SAS. In a comment in your report, identify any problems with the SAS data set that cannot be resolved using the IMPORT procedure. Explain what is causing the problem.
- Read the same raw data file, Pizza.csv, this time using a DATA step (instead of the IMPORT procedure). Be sure to resolve any issues identified above.
- Create a new dataset with the average ratings for each topping.

Answer

a)

The data file Pizza.csv has been uploaded into to SAS using the IMPORT procedure and below are images of the PROC print and PROC contents procedures to review the file contents. The variables are the correct type.

Pizza table PROC print:

Obs	SurveyNum	Arugula	PineNut	Squash	Shrimp	Eggplant
1	101	1	3	3		
2	102	5	4	2		
3	103	4	2	5		
4	104	5	3	2		
5	105	3	5	5		
6	106	2	3	1		
7	107	2	5	5		
8	108	1	3	1		
9	109	4	3	3		
10	110	2	2	5		
11	201	4	2	5		
12	202	3	1	2		
13	203	3	2	5		
14	204	1	5	4		
15	205	3	2	1		
16	206	4	5	1		
17	207	5	3	5		
18	208	1	1	5		
19	209	5	4	3		

Pizza PROC contents report:

The CONTENTS Procedure

Data Set Name	WORK.PIZZA	Observations	120
Member Type	DATA	Variables	6
Engine	V9	Indexes	0
Created	02/04/2022 21:48:23	Observation Length	40
Last Modified	02/04/2022 21:48:23	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	Latin1 Western (Windows)		

Engine/Host Dependent Information

Data Set Page Size	65536
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	1632
Obs in First Data Page	120
Number of Data Set Repairs	0
ExtendObsCounter	YES
Filename	E:\SAS Temporary Files\DXM200017\TD18408_SMVSASCLASSA_\pizza.sas7bdat
Release Created	9.0401M6
Host Created	X64_SRV19
Owner Name	CAMPUS\DXM200017
File Size	128KB
File Size (bytes)	131072

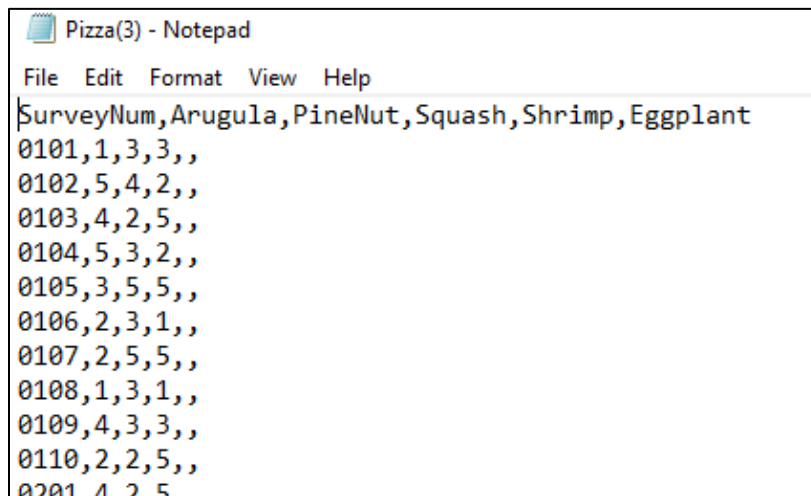
Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
2	Angula	Num	8	BEST12.	BEST32.
6	Eggplant	Char	1	\$1.	\$1.
3	PineNut	Num	8	BEST12.	BEST32.
5	Shrimp	Char	1	\$1.	\$1.
4	Squash	Num	8	BEST12.	BEST32.
1	SurveyNum	Num	8	BEST12.	BEST32.

b)

With PROC Import, SAS automatically scans the first 20 observations to detect the datatype. Missing data is represented by '.' for numeric and blanks for character variables. Since first 20 rows are blanks for variables 'Shrimp' and 'Eggplant', they are identified instead as character variables which is resolved in the below section.

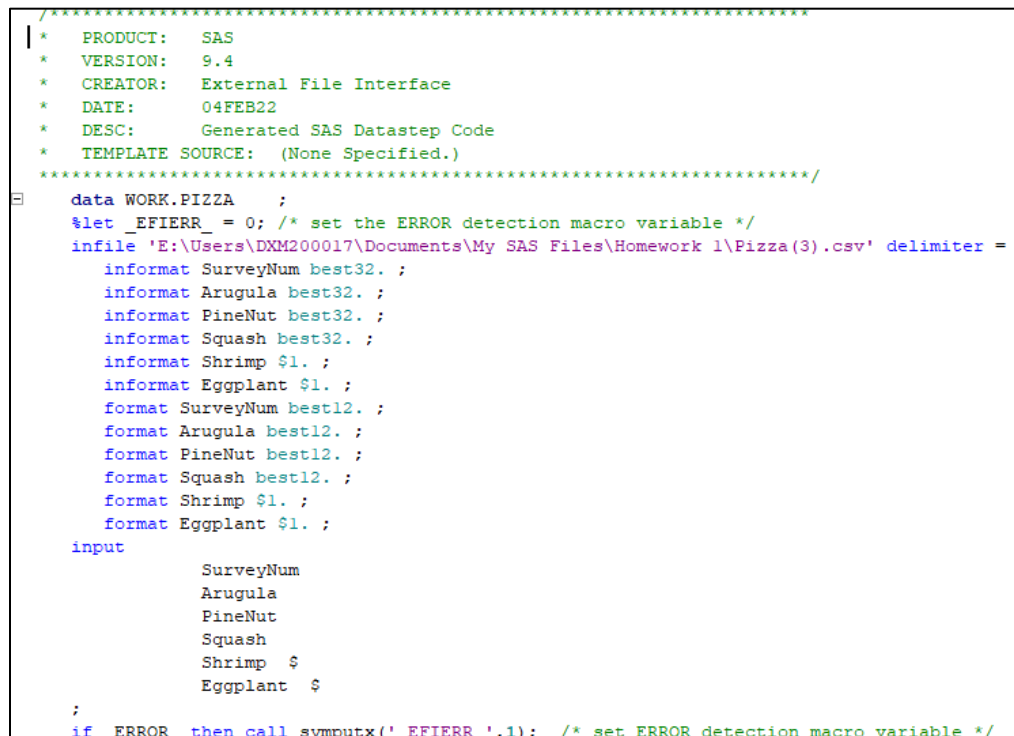
Notepad view of Pizza raw data:



```
Pizza(3) - Notepad
File Edit Format View Help
SurveyNum,Arugula,PineNut,Squash,Shrimp,Eggplant
0101,1,3,3,,
0102,5,4,2,,
0103,4,2,5,,
0104,5,3,2,,
0105,3,5,5,,
0106,2,3,1,,
0107,2,5,5,,
0108,1,3,1,,
0109,4,3,3,,
0110,2,2,5,,
0201,4,2,5,,
```

c)

The DATA step is easily created after using the IMPORT procedure by hitting "F4". See the image below for the automatically created DATA step.



```
*****
* PRODUCT: SAS
* VERSION: 9.4
* CREATOR: External File Interface
* DATE: 04FEB22
* DESC: Generated SAS Datastep Code
* TEMPLATE SOURCE: (None Specified.)
*****/
data WORK.PIZZA ;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'E:\Users\DXM200017\Documents\My SAS Files\Homework 1\Pizza(3).csv' delimiter = ',';
informat SurveyNum best32. ;
informat Arugula best32. ;
informat PineNut best32. ;
informat Squash best32. ;
informat Shrimp $1. ;
informat Eggplant $1. ;
format SurveyNum best12. ;
format Arugula best12. ;
format PineNut best12. ;
format Squash best12. ;
format Shrimp $1. ;
format Eggplant $1. ;
input
    SurveyNum
    Arugula
    PineNut
    Squash
    Shrimp $
    Eggplant $
;
if _ERROR_ then call symputx('_EFIERR_',1); /* set ERROR detection macro variable */
```

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
2	Arugula	Num	8	BEST12.	BEST32.
6	Eggplant	Char	1	\$1.	\$1.
3	PineNut	Num	8	BEST12.	BEST32.
5	Shrimp	Char	1	\$1.	\$1.
4	Squash	Num	8	BEST12.	BEST32.
1	SurveyNum	Num	8	BEST12.	BEST32.

The data step and contents report both showed \$ indicating characters next to the shrimp and eggplant variables. By editing the data step, we can now pull this data into SAS as numbers. See the edited data step below.

```
*****/
data WORK.PIZZA ;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'E:\Users\DXM200017\Documents\My SAS Files\Homework 1\Pizza(3).csv' delimiter = ',' MISSOVER
    informat SurveyNum best32. ;
    informat Arugula best32. ;
    informat PineNut best32. ;
    informat Squash best32. ;
    informat Shrimp best32. ;
    informat Eggplant best32. ;
    format SurveyNum best12. ;
    format Arugula best12. ;
    format PineNut best12. ;
    format Squash best12. ;
    format Shrimp best12. ;
    format Eggplant best12. ;
input
    SurveyNum
    Arugula
    PineNut
    Squash
    Shrimp
    Eggplant
;
if _ERROR_ then call symputx('_EFIERR_',1); /* set ERROR detection macro variable */
run;
```

The “.” in each cell instead of blanks for missing values, indicates that SAS now reads these as numeric values instead of character values.

	SurveyNum	Arugula	PineNut	Squash	Shrimp	Eggplant
1	101	1	3	3	.	.
2	102	5	4	2	.	.
3	103	4	2	5	.	.
4	104	5	3	2	.	.
5	105	3	5	5	.	.
6	106	2	3	1	.	.
7	107	2	5	5	.	.
8	108	1	3	1	.	.
9	109	4	3	3	.	.

d)

The variables were renamed to show the correct names of “Topping” and “Average_Rating”.

Pizza Table:

VIEWTABLE: Work.Pizza_summary		
	Topping	Average_Rating
1	Arugula	3.075000
2	PineNut	3.140000
3	Squash	3.162500
4	Shrimp	2.966667
5	Eggplant	2.860000

Question 2

The new management of a local hotel decided to update their recently acquired (and very outdated) property by installing wireless Internet service for their guests. They are also considering updating their billing system because the method used by the previous owner seems faulty. In order to conduct a billing analysis, they would like some calculations about the guests who stayed with them during the first part of February (this was the first month after the change of ownership). The raw data file Hotel.dat contains variables with information on room number, number of guests, check-in month, day, year, check-out month, day, year, use of wireless Internet service, number of days of Internet use, room type, and room rate.

- Examine the raw data file Hotel.dat and read it into SAS. Next, create date variables for the check-in and check-out dates, and format them to display as readable dates.
- Create a variable that calculates the subtotal as the room rate times the number of days in the stay, plus a per person rate (\$10 per day for each person beyond one guest, for example for 3 guests, the total per person rate will be $(3-1)*10=\$20$), plus an Internet service fee (\$9.95 for a one-time activation and \$4.95 per day of use).
- Create a variable that calculates the grand total as the subtotal plus sales tax at 7.75%. The result should be rounded to two decimal places.
- View the resulting data set. In a comment in your report, state the value for the grand total for room 211.

Answer:

a)

Hotel table with check in and check out dates added and formatted as readable dates:

VIEWTABLE: Work.Hotel														
	roomnumber	numberofguests	chkinmonth	chkinday	chkinyear	chkoutmonth	chkoutday	chkoutyear	wirelessuse	noofdaysinternetuse	roomtype	roomrate	checkindate	checkoutdate
1	211	3	2	7	2014	2	11	2014	NO		Deluxe Suite	295	02/07/2014	02/11/2014
2	214	2	2	2	2014	2	12	2014	NO		Basic no view	75	02/02/2014	02/12/2014
3	216	4	2	2	2014	2	13	2014	NO		Suite	255	02/02/2014	02/13/2014
4	220	5	2	3	2014	2	12	2014	YES		2 Basic w/view	155	02/03/2014	02/12/2014
5	221	3	2	3	2014	2	12	2014	NO		Luxury	195	02/03/2014	02/12/2014
6	223	5	2	7	2014	2	13	2014	NO		Suite	255	02/07/2014	02/13/2014

roomrate	checkindate	checkoutdate
295	02/07/2014	02/11/2014
75	02/02/2014	02/12/2014
255	02/02/2014	02/13/2014
155	02/03/2014	02/12/2014
195	02/03/2014	02/12/2014
255	02/07/2014	02/13/2014

b)

We made the assumption that the one-time activation fee is only charged when a guest uses the wireless internet service. A number of days variable “noofdays” was included to complete the daily rate calculation.

Hotel table with number of days and subtotal added:

VIEWTABLE: Work.Hotel						
	roomtype	roomrate	checkindate	checkoutdate	noofdays	subtotal
1	Deluxe Suite	295	02/07/2014	02/11/2014	4	1260
2	Basic no view	75	02/02/2014	02/12/2014	10	850
3	Suite	255	02/02/2014	02/13/2014	11	3135
4	Basic w/view	155	02/03/2014	02/12/2014	9	1774.85
5	Luxury	195	02/03/2014	02/12/2014	9	1935
6	Suite	255	02/07/2014	02/13/2014	6	1770
7	Basic w/view	155	01/31/2014	02/13/2014	13	2464.45

c)

Hotel table with grandtotal added:

VIEWTABLE: Work.Hotel							
	roomtype	roomrate	checkindate	checkoutdate	noofdays	subtotal	grandtotal
1	Deluxe Suite	295	02/07/2014	02/11/2014	4	1260	1357.65
2	Basic no view	75	02/02/2014	02/12/2014	10	850	915.88
3	Suite	255	02/02/2014	02/13/2014	11	3135	3377.96
4	Basic w/view	155	02/03/2014	02/12/2014	9	1774.85	1912.4
5	Luxury	195	02/03/2014	02/12/2014	9	1935	2084.96
6	Suite	255	02/07/2014	02/13/2014	6	1770	1907.18
7	Basic w/view	155	01/31/2014	02/13/2014	13	2464.45	2655.44
8	Luxury	195	02/01/2014	02/13/2014	12	2364.8	2548.07
9	Deluxe Suite	295	02/03/2014	02/12/2014	9	3069.5	3307.39

d)

The grand total for room 211 is 1357.65.

PROC Print of final Hotel table:

The SAS System																	
Obs	roomnumber	numberofguests	chkinmonth	chkinday	chkinyear	chkoutmonth	chkoutday	chkoutyear	wirelessuse	noofdaysinternetuse	roomtype	roomrate	checkindate	checkoutdate	noofdays	subtotal	grandtotal
1	211	3	2	7	2014	2	11	2014	NO	.	Deluxe Suite	295	02/07/2014	02/11/2014	4	1260.00	1357.65

roomtype	roomrate	checkindate	checkoutdate	noofdays	subtotal	grandtotal
Deluxe Suite	295	02/07/2014	02/11/2014	4	1260.00	1357.65