1. B.

Model	Price	CustNumber	Quantity	TotalCost
L776	$159.98	101	1	$159.98
M123	$4.59	102	10	$45.90
M567	$23.50	103	1	$23.50
X999	$29.95	103	2	$59.90

1. D.

Model	Price
S776	$1.99
S888	$12.99

1. E.

Included is the log from running the data step. I have not included the note regarding the original data being in a format native to another host, as that is not relevant.

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
72
73         data hw4.purchase_price_joshlj2
74          hw4.not_purchased_joshlj2(KEEP = Model
75            Price);
76          merge hw4.INVENTORY(in=inInv)
77             hw4.PURCHASE(in=inPur);
78          by Model;
79          if inInv and inPur then do;
80          TotalCost = Quantity*Price;
81          format TotalCost DOLLAR8.2;
82          output hw4.purchase_price_joshlj2;
83          end;
84
85          if not inPur;
86          output hw4.not_purchased_joshlj2;
87         run;

NOTE: There were 6 observations read from the data set HW4.INVENTORY.
NOTE: There were 4 observations read from the data set HW4.PURCHASE.
NOTE: The data set HW4.PURCHASE_PRICE_JOSHLJ2 has 4 observations and 5 variables.
NOTE: The data set HW4.NOT_PURCHASED_JOSHLJ2 has 2 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.08 seconds
      cpu time            0.03 seconds
```

```
88
89
90          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
102
```

2. A.

Dataset	Total Variables	Total Observations
FMLI141	30	1734
FMLI142	30	3319
FMLI143	30	4879
FMLI144	30	6280
MEMI141	9	4251
MEMI142	9	8252
MEMI143	9	12032
MEMI144	9	15512

2. C.

The CONTENTS Procedure

Data Set Name	HW4.FMLI2014_JOSHLJ2	**Observations**	16212
Member Type	DATA	**Variables**	31
Engine	V9	**Indexes**	0
Created	10/30/2019 20:13:22	**Observation Length**	152
Last Modified	10/30/2019 20:13:22	**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)		

The results above are what I expect. We should have a total of 1734 + 3319 + 4879 + 6280 = 16212 observations, which we see. We also see that there are 31 variables rather than 30, since we have added a quarter variable. This description of the dataset looks good to me.

2. E.

The CONTENTS Procedure

Data Set Name	HW4.MEMI2014_JOSHLJ2	**Observations**	40047
Member Type	DATA	**Variables**	10
Engine	V9	**Indexes**	0
Created	10/30/2019 20:13:22	**Observation Length**	48
Last Modified	10/30/2019 20:13:22	**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)		

The results above are what I expect. We should have a total of 4251 + 8252 + 12032 + 15512 = 40047 observations, which we see. We also see that there are 10 variables rather than 9, since we have added a quarter variable. This description of the dataset looks good to me.

2. F.

The FREQ Procedure

SALARYX	Frequency	Percent
Frequency Missing = 27674		
Impoverished	2672	21.60
Lower Class	5104	41.25
Middle Class	3249	26.26
Upper Middle Class	746	6.03
Upper Class	602	4.87

The table above is a frequency table on the SALARYX variable that has been formatted. We can see that there are 27,674 missing observations. The highest salary class is lower class, at 41.25% of the data while the lowest salary class is at 4.87% of the data. This tells us that a majority individuals taking these interviews are considered impoverished or low class. Overall, this table gives a good idea of missing data as well as where most interviewees lie economic class-wise.

The FREQ Procedure

Table of SALARYX by QTR

SALARYX	QTR				
	1	2	3	4	Total
Impoverished	279	541	822	1030	2672
	2.25	4.37	6.64	8.32	21.60
Lower Class	560	1077	1560	1907	5104
	4.53	8.70	12.61	15.41	41.25
Middle Class	348	652	976	1273	3249
	2.81	5.27	7.89	10.29	26.26
Upper Middle Class	68	149	220	309	746
	0.55	1.20	1.78	2.50	6.03
Upper Class	53	134	176	239	602
	0.43	1.08	1.42	1.93	4.87
Total	1308	2553	3754	4758	12373
	10.57	20.63	30.34	38.45	100.00

Frequency Missing = 27674

Frequency
Percent

The table above shows us a frequency table between SALARYX and our created QTR variable. For the quarter variable, a majority of our observations are in the 3rd and 4th quarter at about 70% of the data. The 1st and 2nd quarter make up about 30% of the data. This remains true throughout the table between salary and quarter, where the 3rd and 4th quarter are most popular as well as impoverished, lower, and middle class. Overall, this table shows us that most salary frequency remains stable between quarters.

2. H.

The CONTENTS Procedure

Data Set Name	HW4.CE2014_JOSHLJ2	**Observations**	40047
Member Type	DATA	**Variables**	39
Engine	V9	**Indexes**	0
Created	10/30/2019 20:13:23	**Observation Length**	184
Last Modified	10/30/2019 20:13:23	**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)		

The results above are expected. We see that we have 39 variables and 40047 observations. We have 39 variables since our above datasets had 31 and 10 variables. Since they share common variables NEWID and QTR, we can drop the duplicates leaving us with 41-2=39 variables. Observation wise, it makes sense to have 40047 observations. This is the same number of observations as the memi2014 dataset. Since we are merging on NEWID, each observation in FMLI2014 has a unique NEWID. Each observation in memi2014 has a NEWID per family member. So, we are doing a one-to-many merge which results to having the same number of observations as memi2014. Overall, these results make sense observation and variable wise.

2. J.

The CONTENTS Procedure

Data Set Name	HW4.ATLEAST_THREE_JOSHLJ2	**Observations**	2503
Member Type	DATA	**Variables**	1
Engine	V9	**Indexes**	0
Created	10/30/2019 20:13:23	**Observation Length**	160
Last Modified	10/30/2019 20:13:23	**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)		

The CONTENTS Procedure

Data Set Name	HW4.ALL_FOUR_JOSHLJ2	**Observations**	1005
Member Type	DATA	**Variables**	1
Engine	V9	**Indexes**	0
Created	10/30/2019 20:13:23	**Observation Length**	160
Last Modified	10/30/2019 20:13:23	**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)		