Questions:

1. Read the data into SAS. Report what you observe about the datasets (10 pts)

Ans) The dataset salesdata.csv is composed of 300 observations, the variables used here are as follows: internetadvt, price, printadvt, radioadvt , sales, week. In total we have 6 variables. The variable type used here is Num and Len 8.

Engine/Host Dependent Information	
Data Set Page Size	65536
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	1361
Obs in First Data Page	300
Number of Data Set Repairs	0
ExtendObsCounter	YES
Filename	C:\Users\SMUKHE~1\AppData\Local\Temp\SAS Temporary Files_TD3564_SSB401H11_\salesdata.sas7bdat
Release Created	9.0401M6
Host Created	X64_10PRO
Owner Name	ILTECH\smukherjee3
File Size	128KB
File Size (bytes)	131072

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
5	internetadvt	Num	8	BEST12.	BEST32.
2	price	Num	8	BEST12.	BEST32.
3	printadvt	Num	8	BEST12.	BEST32.
4	radioadvt	Num	8	BEST12.	BEST32.
1	sales	Num	8	BEST12.	BEST32.
6	week	Num	8	BEST12.	BEST32.

DATASET STRUCTURE

The SAS System

Obs	sales	price	printadvt	radioadvt	internetadvt	week
1	59	3.08	12	9	6	1
2	99	6.18	11	12	16	2
3	67	5.87	12	11	8	3
4	99	6.4	14	12	15	4
5	85	6.04	11	13	12	5
6	67	4.74	13	10	7	6
7	83	2.85	14	10	10	7
8	45	6.44	11	6	6	8
9	72	3.14	11	7	11	9
10	115	5.93	14	9	20	10
11	94	3.76	11	8	16	11
12	80	3.62	13	8	11	12
13	106	4.56	12	12	16	13
14	107	3.69	13	6	20	14
15	80	5.76	11	11	11	15
16	99	4.59	12	7	18	16
17	96	3.99	15	12	13	17
18	72	5.95	10	10	11	18
19	92	2.75	12	10	13	19
20	64	3.63	14	10	6	20

1ST 20 OBSERVATIONS SHOWN ABOVE

2. Study the raw data and report what you see from the data. (10 pts)

The SAS System

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
sales	300	83.803	18.592	45.000	125.000
price	300	4.537	1.415	2.000	6.970
printadvt	300	12.350	1.400	10.000	15.000
radioadvt	300	9.053	2.360	5.000	13.000
internetadvt	300	12.437	4.244	5.000	20.000
week	300	150.500	86.747	1.000	300.000

Summary statistics of the dataset.

Explanation:

The dataset comprises of 300 observations and revealing a certain degree of variability across the listed marketing metrics. On an average sales hover around an average of 83 UNITS, with a wide range between 45 UNITS to 125 UNITS. Average pricing is about 4.5 dollars with a variability of 2 to 7 dollars approx. which is significant. Average cost involved for printadvt is is 12.350, in a range of 10 to 15 dollars of variability.Average cost of radioadvyt is 9.05 approx lying in a range of 5 to 13 dollars and internetadvt of average of 124.437 dollars with a wide variability of 5 dollars to 20 dollars.

Note weeks for which we are assessing the model are between 1 – 300 too.

3. Propose a suitable model to analyze this data. Specify the model specification. Setup the data for analysis and run the analysis using SAS. Describe and justify all your modeling decisions. (10 pts)

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: sales

Number of Observations Read	300
Number of Observations Used	300

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1556.90400	1556.90400	4.56	0.0336
Error	298	101800	341.61239		
Corrected Total	299	103357			

Root MSE	18.48276	R-Square	0.0151
Dependent Mean	83.80333	Adj R-Sq	0.0118
Coeff Var	22.05492		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	91.11777	3.58856	25.39	<.0001
price	1	-1.61219	0.75518	-2.13	0.0336

The SAS System

Obs	_MODEL_	_TYPE_	_DEPVAR_	_RMSE_	Intercept	price	sales
1	MODEL1	PARMS	sales	18.4828	91.1178	-1.61219	-1

The regression denotes that the coefficient of price is negative (-1.61), suggesting higher prices reduce sales. P value pf 0.0336. , and the model explains almost none of the variation in sales (adjusted r square of = 0.03, which indicates that the price of the product is not strongly driven by price alone and other facrs such as advertising play a bigger role.

Key variables :

Intercept:91.117

Priceperunit = -1.162

Sales = 91.117 –1.162 * PRICE

NOW I AM RUNNING REGRESSION ON PRICE + PRINTADVT+RADIOADVT+INTERNETADVT + **WEEK(NOT A PREDICTOR)**

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: sales

Number of Observations Read	300
Number of Observations Used	300

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2783.54883	1391.77442	4.11	0.0174
Error	297	100574	338.63248		
Corrected Total	299	103357			

Root MSE	18.40197	R-Square	0.0269
Dependent Mean	83.80333	Adj R-Sq	0.0204
Coeff Var	21.95852		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	72.93088	10.20183	7.15	<.0001
price	1	-1.54537	0.75270	-2.05	0.0409
printadvt	1	1.44807	0.76084	1.90	0.0580

The SAS System

The REG Procedure
Model: MODEL4
Dependent Variable: sales

Number of Observations Read	300
Number of Observations Used	300

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	14361	7180.74282	23.96	<.0001
Error	297	88996	299.64953		
Corrected Total	299	103357			

Root MSE	17.31039	R-Square	0.1389
Dependent Mean	83.80333	Adj R-Sq	0.1332
Coeff Var	20.65597		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	67.04294	4.98593	13.45	<.0001
price	1	-1.84640	0.70819	-2.61	0.0096
radioadvt	1	2.77660	0.42475	6.54	<.0001

When I regressed the price variable against advertising factors such as printadvt and advt (the effects were statistically significant) as the p values of printadvt and radioadvt clearly tells us both are lesser than 0.5, with an adjusted r square of 0.0204 and 0.1332,which means printadvt doesn't impact the price as much as radioadvt does.

p.t.o

The SAS System

The REG Procedure
Model: MODEL2
Dependent Variable: sales

Number of Observations Read	300
Number of Observations Used	300

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	89815	44907	984.84	<.0001
Error	297	13543	45.59861		
Corrected Total	299	103357			

Root MSE	6.75267	R-Square	0.8690
Dependent Mean	83.80333	Adj R-Sq	0.8681
Coeff Var	8.05776		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	41.94858	1.72279	24.35	<.0001
price	1	-1.87390	0.27597	-6.79	<.0001
internetadvt	1	4.04904	0.09203	43.99	<.0001

when I incorporated the internetadvt variable , we see a result of adjusted r square of 0.8681,that explains massive variation in sales, and impacts the price the most.

4. Based on your model estimates, what can you tell Mr. Bond about advertising channels used for his product? (10 pts)

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: sales

R-Square Selection Method

Number of Observations Read	300
Number of Observations Used	300

Number in Model	R-Square	Parameter Estimates					
		Intercept	price	printadvt	radioadvt	internetadvt	week
1	0.8486	33.61429	.	.	.	4.03557	.
2	0.9483	11.64639	.	.	2.48806	3.99075	.
3	0.9734	20.37301	-2.08618	.	2.55056	4.00463	.
4	0.9867	1.45048	-2.00994	1.54037	2.47459	4.02399	.
5	0.9867	1.43619	-2.00971	1.54030	2.47441	4.02403	0.00010190

The SAS System

Obs	_MODEL_	_TYPE_	_DEPVAR_	_RMSE_	Intercept	price	printadvt	radioadvt	internetadvt	week	sales	_IN_	_P_	_EDF_	_RSQ_
1	MODEL1	PARMS	sales	7.24574	33.6143	.	.	.	4.03557	.	-1	1	2	298	0.84863
2	MODEL1	PARMS	sales	4.24344	11.6464	.	.	2.48806	3.99075	.	-1	2	3	297	0.94826
3	MODEL1	PARMS	sales	3.04735	20.3730	-2.08618	.	2.55056	4.00463	.	-1	3	4	296	0.97341
4	MODEL1	PARMS	sales	2.15667	1.4505	-2.00994	1.54037	2.47459	4.02399	.	-1	4	5	295	0.98672
5	MODEL1	PARMS	sales	2.16031	1.4362	-2.00971	1.54030	2.47441	4.02403	.000101905	-1	5	6	294	0.98672

A stepwise modeling approach was used to identify the optimal specification, starting with the simpler models including single advertising variables and additional variables which were incrementally added based on their contribution to explaining sales variation**. I witnessed that when regressed price against just internetadvt (advertising variable) is the variable having the highest adjusted rsquare value 0.8681 and r square of 0.8691 hence this could have an effect on the price of the product.**

The SAS System

The MEANS Procedure

Variable	Mean
sales	83.8033333
price	4.5369667
printadvt	12.3500000
radioadvt	9.0533333
internetadvt	12.4366667
week	150.5000000

The SAS System

Obs	_TYPE_	_FREQ_	sales_Mean	price_Mean	printadvt_Mean	radioadvt_Mean	internetadvt_Mean	week_Mean
1	0	300	83.803333333	4.5369666667	12.35	9.0533333333	12.436666667	150.5

5. What would happen to sales if Mr. Bond increased the advertising in each of the channels? (10 pts)

The SAS System

Obs	_MODEL_	_DEPVAR_	_RMSE_	Intercept	price	sales	_FREQ_	sales_Mean	price_Mean	printadvt_Mean	radioadvt_Mean	internetadvt_Mean	week_Mean	E_P_price
1	MODEL1	sales	18.4828	91.1178	-1.61219	-1	300	83.803333333	4.5369666667	12.35	9.0533333333	12.436666667	150.5	-0.087281

FROM A MARKETING ANALYST'S POINT OF VIEW, THE DATA SHOWS THAT PRODUCT DEMAND IS QUITE INELASTIC WHICH MEANS MOST CUSTOMERS AREN'T TOO SENSITIVE TO PRICE CHANGES, WHILE THE REGRESSION CONFIRMS THAT SALES DO DIP SLIGHTLY WHEN PRICES GO UP (WITH A COEFFICIENT OF – 1.61), THE REAL TAKEAWAY COMES FROM THE PRICE ELASTICITY OF DEMAND , WHICH IS JUST -0.08.IN SIMPLE TERMS IF THE PRICES FOR THE ADVERTISING CHANNELS WERE TO GO UP/ PRICE OF THE PRODUCT WOULD GO UP BY 1 %, SALES WOULD DROP BY ONLY 0.087.THIS EXPLAINS THE CONSUMERS ARE MUCH LIKELY TO KEEP BUYING THE PRODUCTS EVEN IF THE PRICES RISE A BIT. WHICH MEANS ,YHE COMPANY COULD CONSIDER MODESTLY INCREASING THE PRICES OF THE PRODUCT BY INCREASING ADVERTISING WITHOUT HURTING SALES.THIS COYULD IMPROVE OVERALL REVENUE.

APPENDIX:

```
libname exam "C:\Users\smukherjee3\Downloads\final";
proc import out = salesdata
datafile = "C:\Users\smukherjee3\Downloads\final\salesdata.csv"
dbms = csv replace;
getnames = yes;
datarow = 2;
run;
*printing 1st 10 observations;
proc print data= salesdata (obs=20);
run;
*viewing the dataset structure and initial observations;
proc contents data= salesdata;
run;
*creating summarys statistics;
proc means data= salesdata maxdec=3;
run;
*print regression results results to review impact of price on sales and omarketing variables;
proc reg data= salesdata plots=none outest =linreg_adv;
model sales = price;
run;quit;
proc print data = linreg_adv;
run;

proc reg data= salesdata plots=none;
model sales = price;
model sales = price internetadvt;
```

```
model sales = price printadvt;
model sales = price radioadvt;
*model sales = price printadvt radioadvt;
*model sales = price printadvt radioadv internetadvt;
*model sales = price printadvt radioadv internetadvt week;
run;quit;
*automated model selection;
proc reg data = salesdata plots=none outest=linreg_advert;
model sales = price printadvt radioadvt internetadvt week / selection=rsquare b best=1;
run;

proc print data=linreg_advert;
run;
*Calculating mean values for elasticity analysis;
proc means data = salesdata mean;
var sales price printadvt radioadvt internetadvt week;
output out=means_all_v mean= /autoname;
run;

proc print data = means_all_v;
run;
        *merge regression estimates and mean values for elasticity calculations;
data elasticity;
  merge linreg_adv(drop=_TYPE_)
       means_all_v(drop=_TYPE_);
  E_P_price = price * (price_Mean / sales_Mean);
run;
```

```
proc print data=elasticity;
run;
```