Data Set Name	WORK.DIAMONDS	Observations	53940
Member Type	DATA	Variables	11
Engine	V9	Indexes	0
Created	18/01/2023 10:58:10	Observation Length	80
Last Modified	18/01/2023 10:58:10	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	34
First Data Page	1
Max Obs per Page	1635
Obs in First Data Page	1595
Number of Data Set Repairs	0
Filename	/saswork/SAS_work68D500013F65_odaws01-usw2-2.oda.sas.com/SAS_work26FA00013F65_odaws01-usw2-2.oda.sas.com/diamonds.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	671088824
Access Permission	rw-r--r--
Owner Name	u62797381
File Size	4MB
File Size (bytes)	4587520

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
1	VAR1	Char	4	$4.	$4.
2	carat	Num	8	BEST12.	BEST32.
5	clarity	Char	6	$6.	$6.
4	color	Char	3	$3.	$3.
3	cut	Char	11	$11.	$11.
6	depth	Num	8	BEST12.	BEST32.
8	price	Num	8	BEST12.	BEST32.
7	table	Num	8	BEST12.	BEST32.
9	x	Num	8	BEST12.	BEST32.
10	y	Num	8	BEST12.	BEST32.
11	z	Num	8	BEST12.	BEST32.

Variable	Mean	Std Dev	Minimum	Maximum	N
y	5.7345260	1.1421347	0	58.9000000	53940
z	3.5387338	0.7056988	0	31.8000000	53940
carat	0.7979397	0.4740112	0.2000000	5.0100000	53940
depth	61.7494049	1.4326213	43.0000000	79.0000000	53940
table	57.4571839	2.2344906	43.0000000	95.0000000	53940
price	3932.80	3989.44	326.0000000	18823.00	53940
x	5.7311572	1.1217607	0	10.7400000	53940

Distribution of y

Distribution of z

Distribution of y

Distribution of z

Which color is prefered the most and which one is the least?

Which clarity is prefered the most?

Which cut is prefered the most and least?

Which carat is considered better?

Pairplot
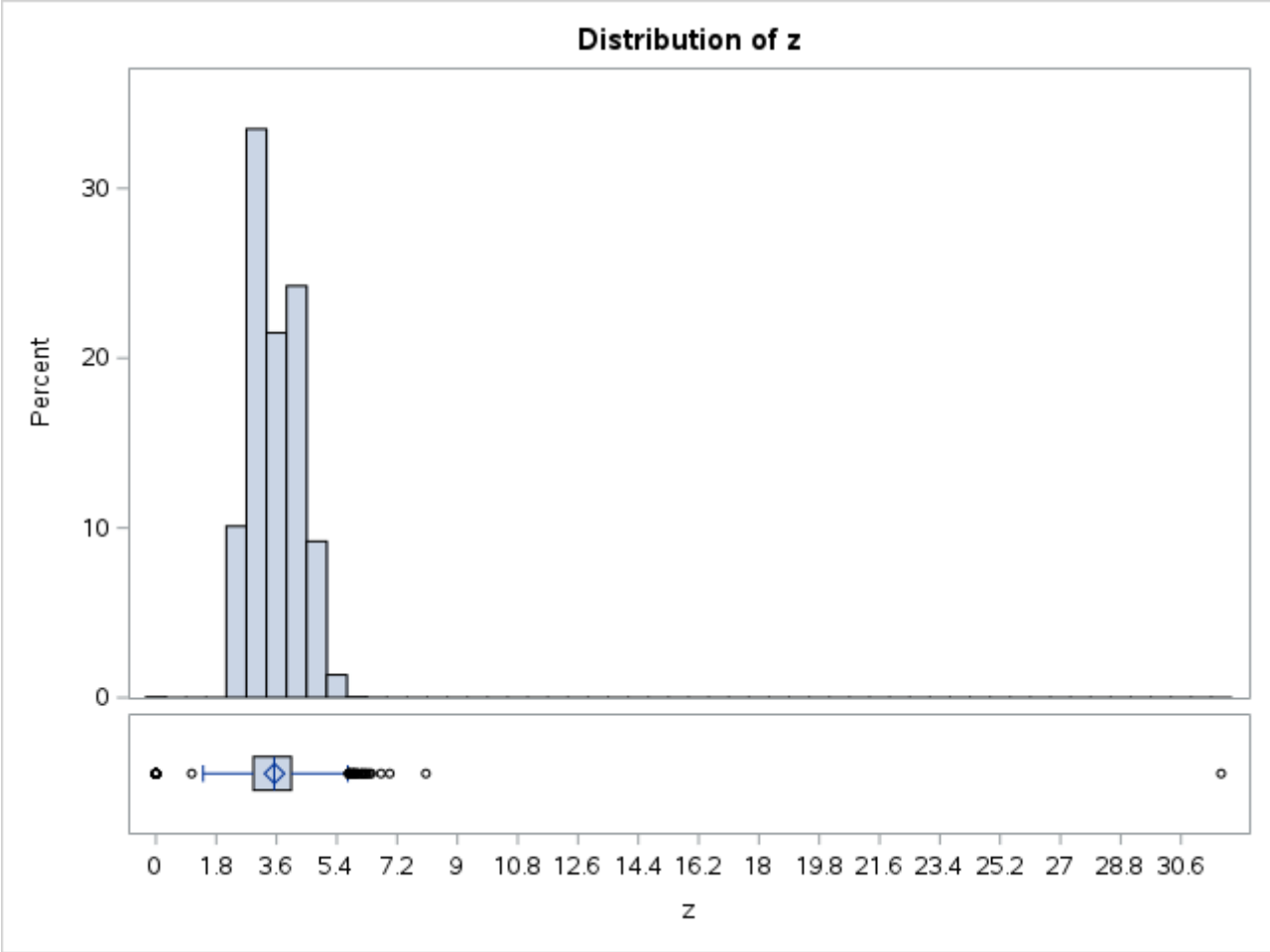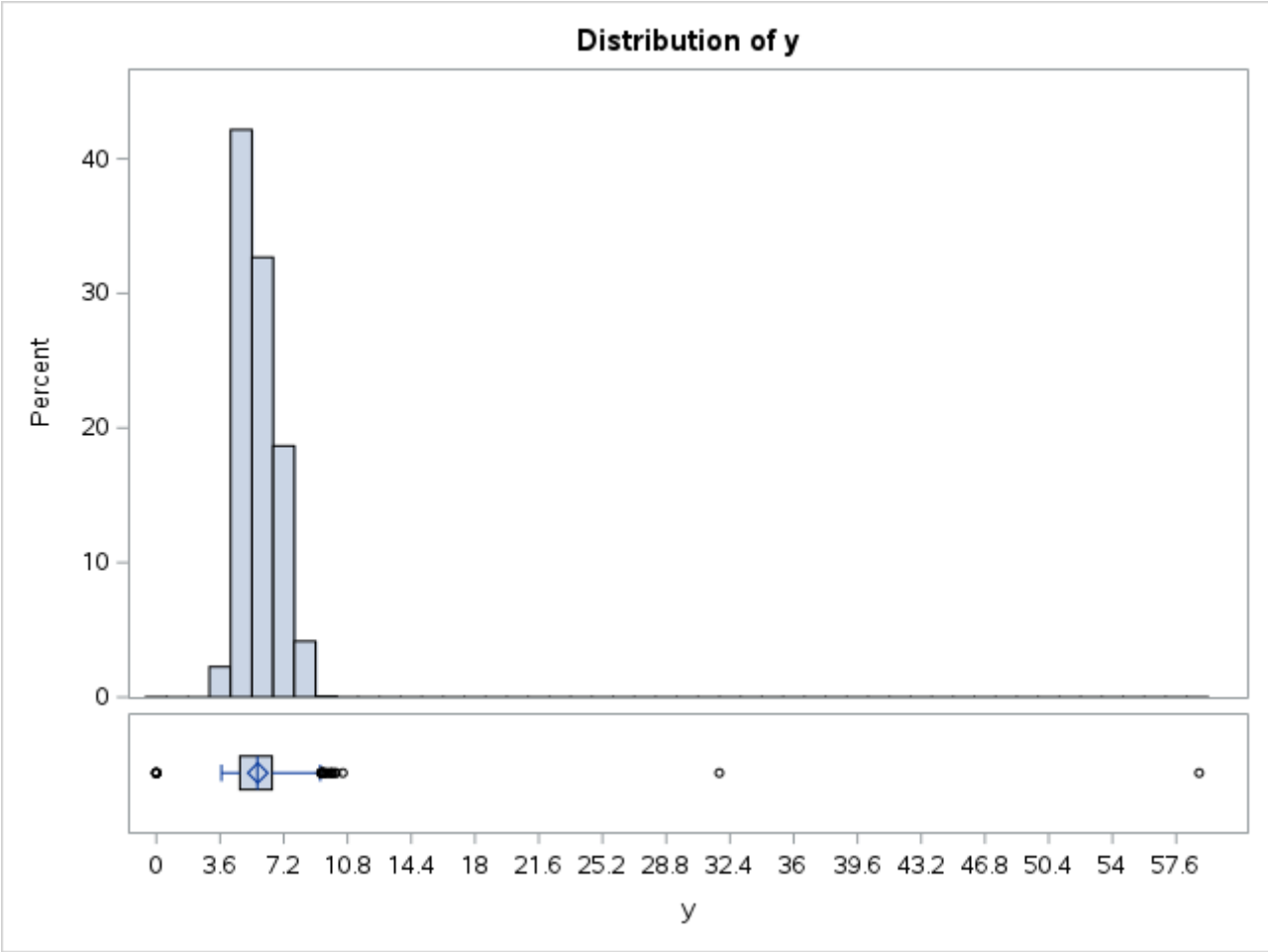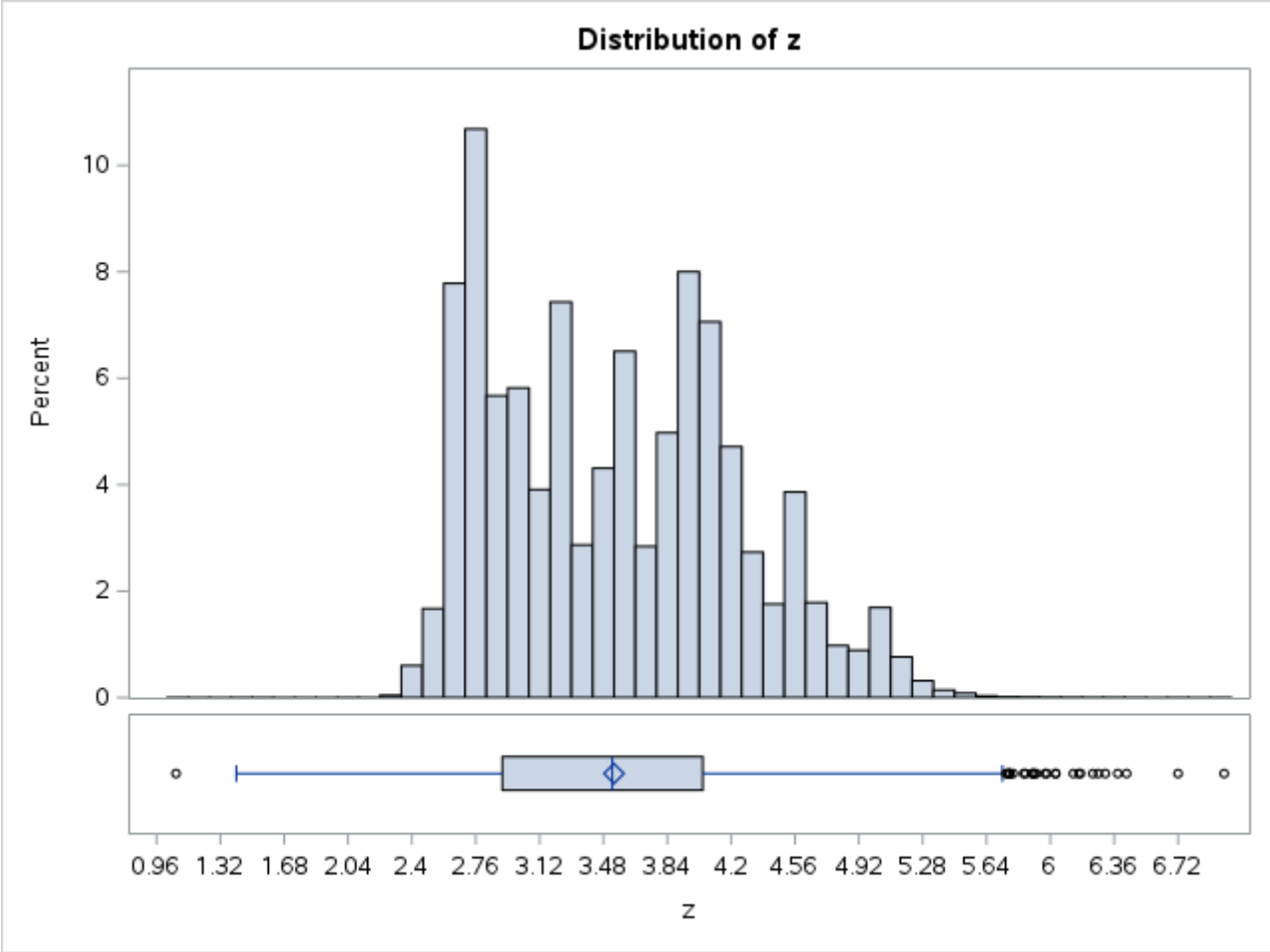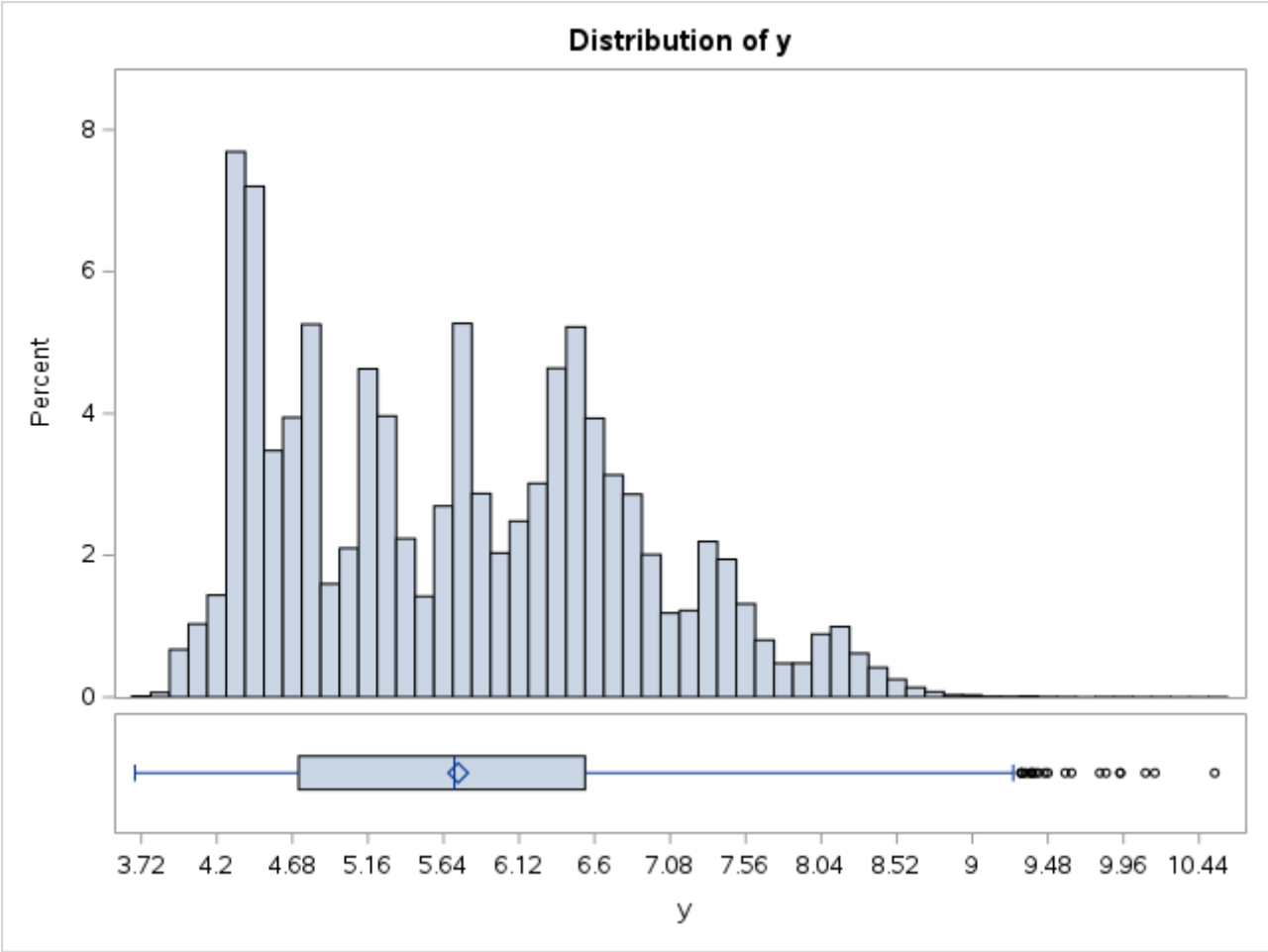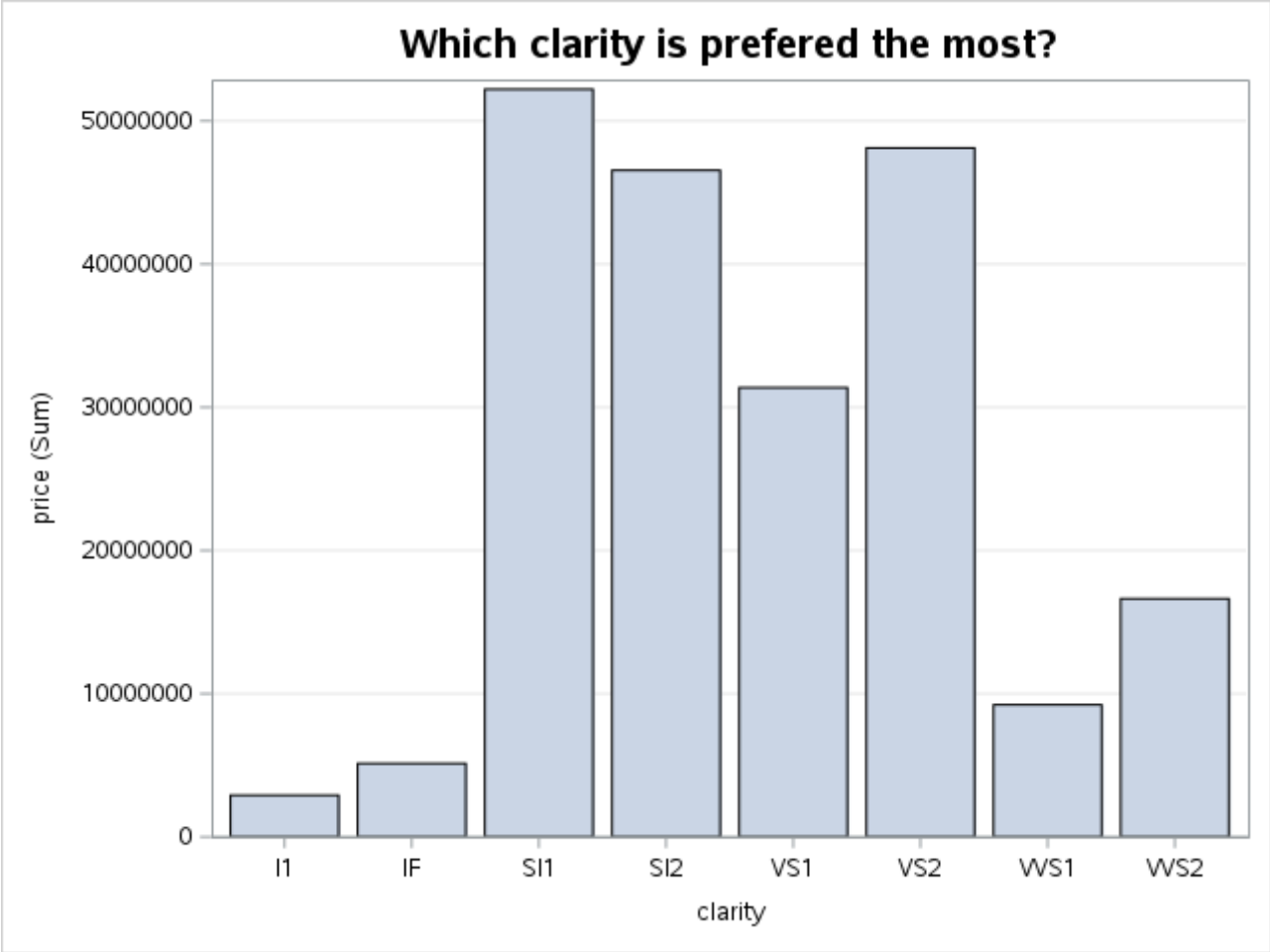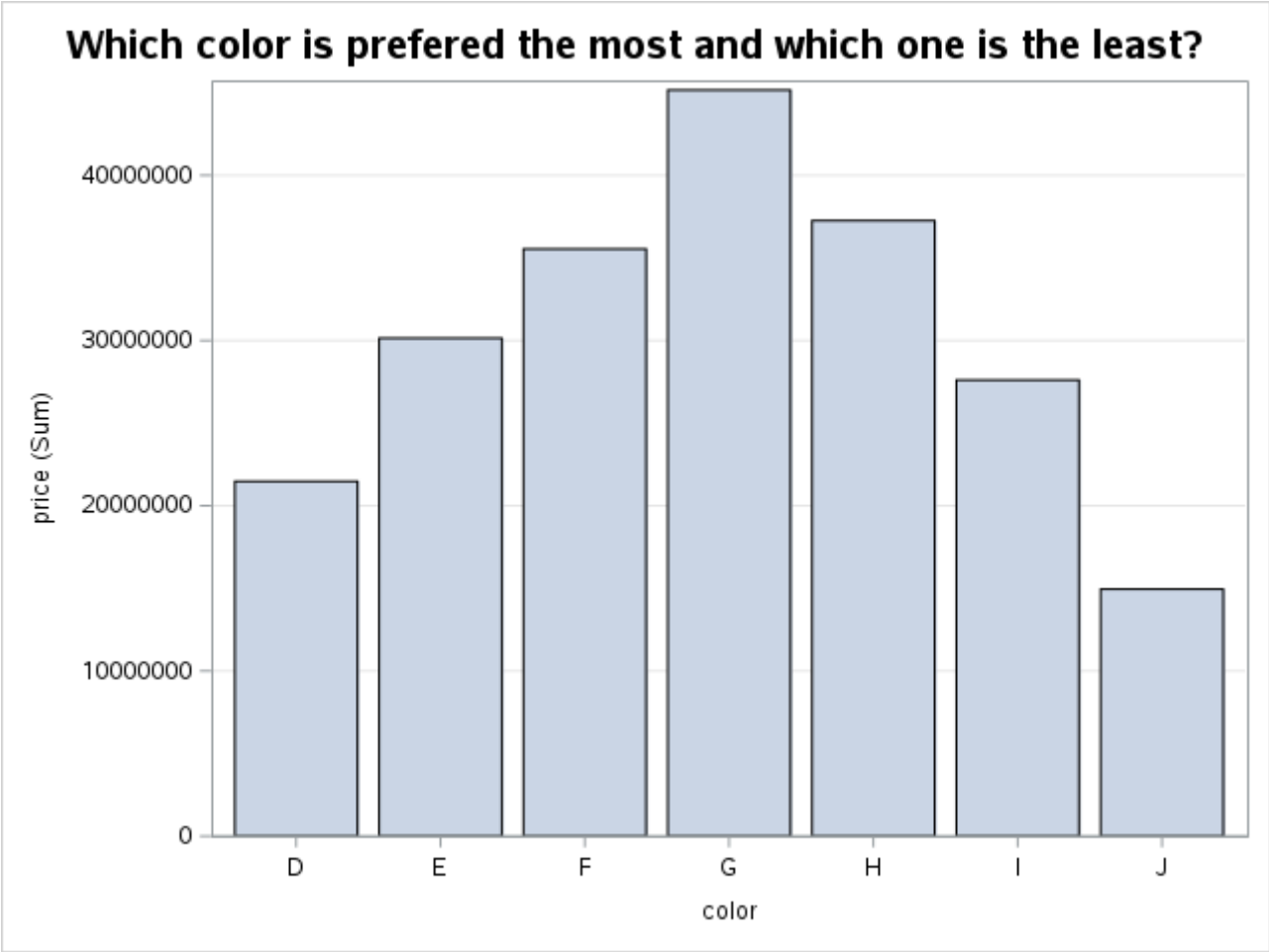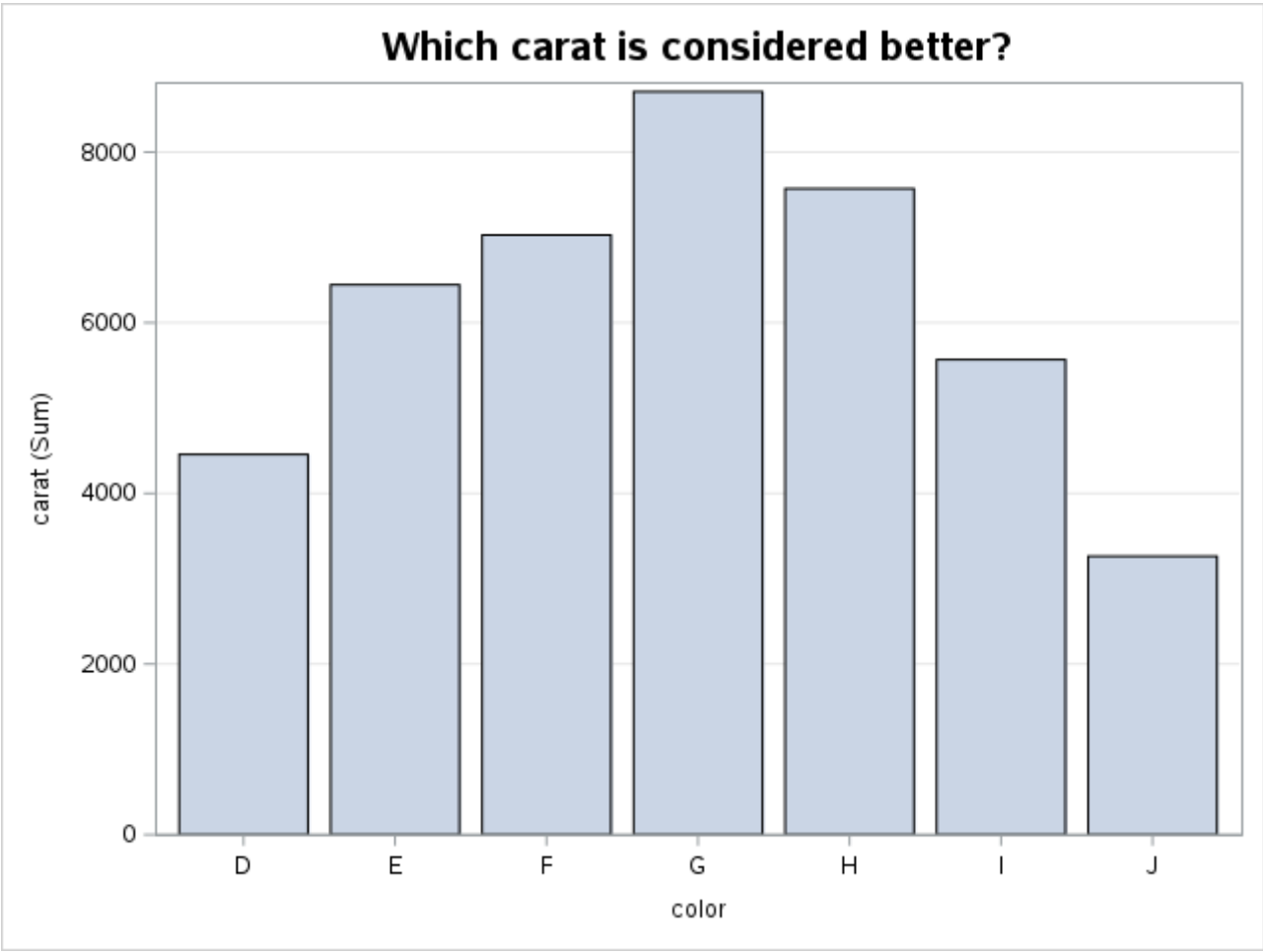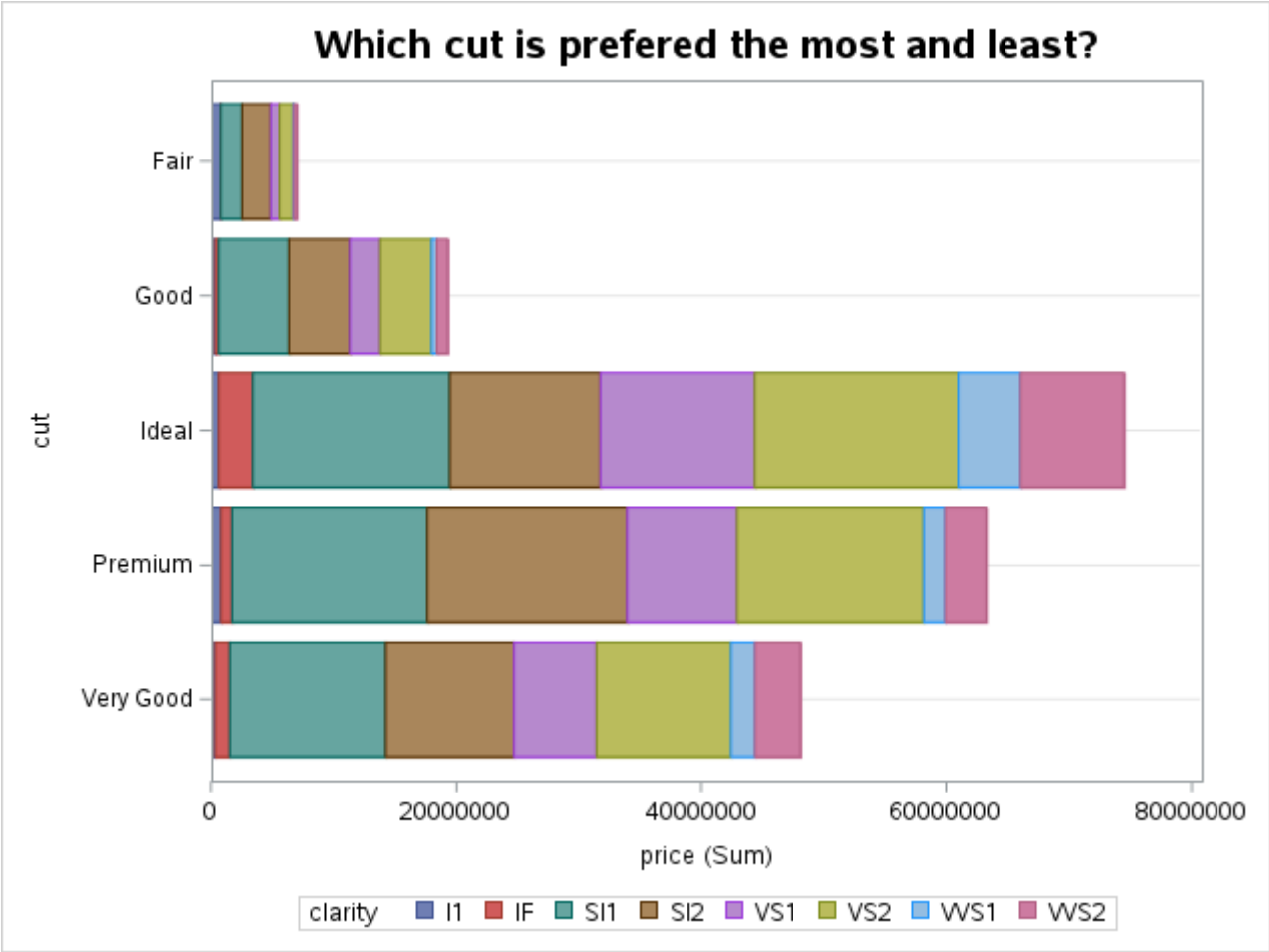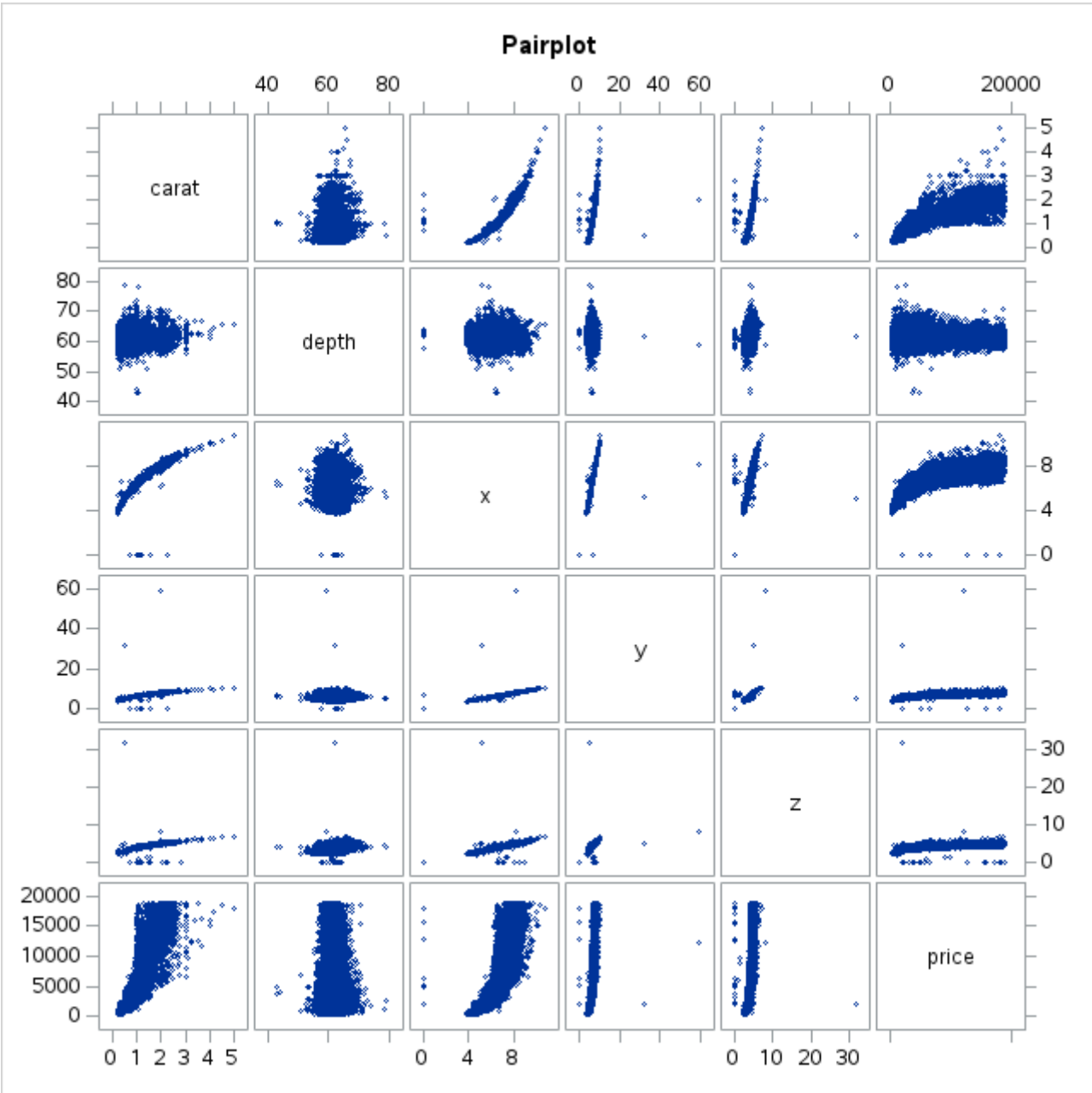
6 With Variables:	carat depth table x y z
1 Variables:	price

Pearson Correlation Coefficients, N = 53940	
	price
carat	0.92159
depth	-0.01065
table	0.12713
x	0.88444
y	0.86542
z	0.86125

Data Set	WORK.DIAMOND_LOG
Factor Extraction Method	Partial Least Squares
PLS Algorithm	NIPALS
Number of Response Variables	1
Number of Predictor Parameters	1

Missing Value Handling	Exclude
Number of Factors	1

Number of Observations Read	53917
Number of Observations Used	53917

Percent Variation Accounted for by Partial Least Squares Factors				
	Model Effects		Dependent Variables	
Number of Extracted Factors	Current	Total	Current	Total
1	100.0000	100.0000	84.9320	84.9320

Fit Plot for carat

Data Set	WORK.DIAMOND_LOG
Factor Extraction Method	Partial Least Squares
PLS Algorithm	NIPALS
Number of Response Variables	1
Number of Predictor Parameters	1
Missing Value Handling	Exclude
Number of Factors	1

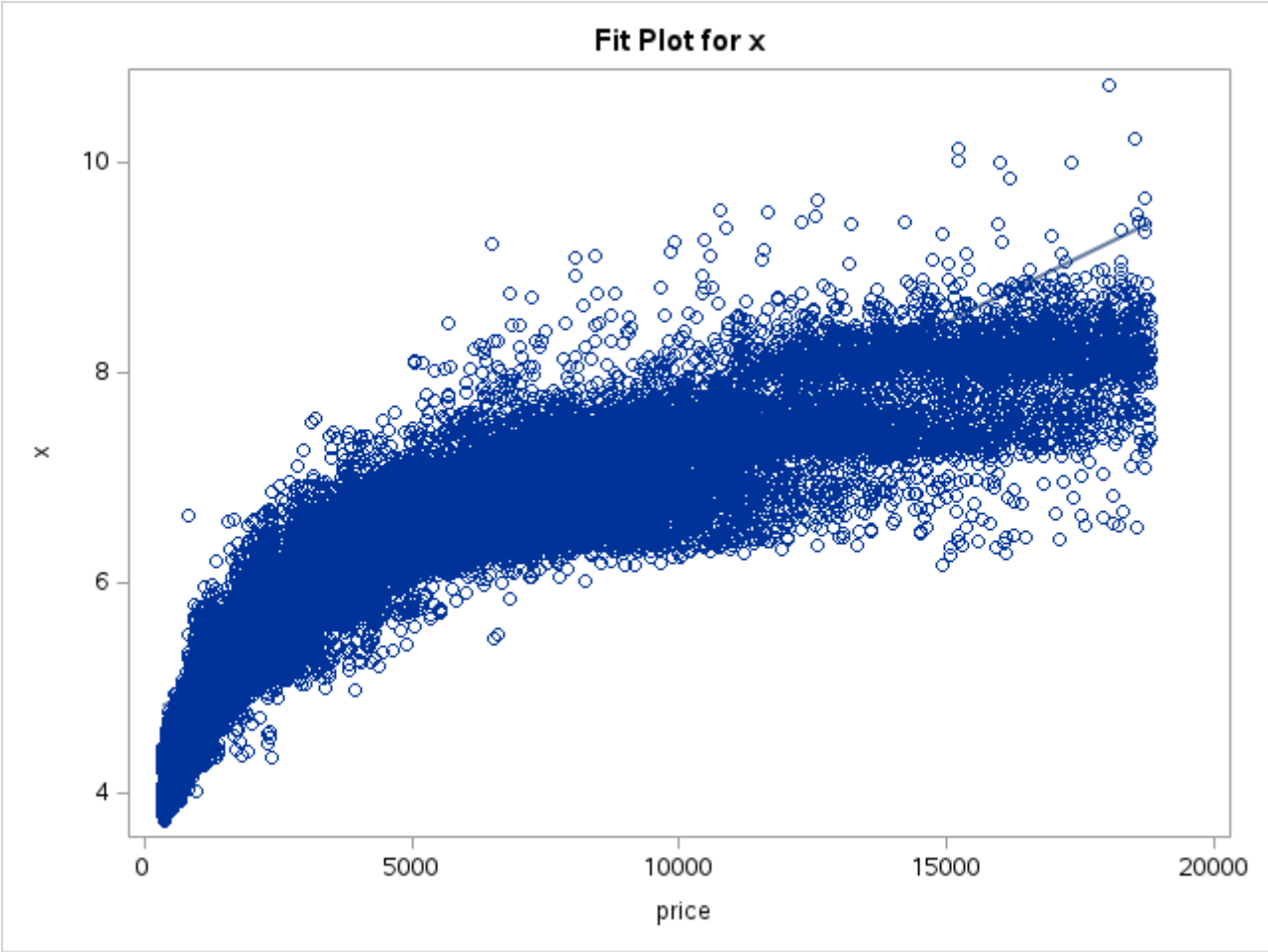| Number of Observations Read	53917
Number of Observations Used	53917

Percent Variation Accounted for by Partial Least Squares Factors

Number of Extracted Factors	Model Effects		Dependent Variables	
	Current	Total	Current	Total
1	100.0000	100.0000	78.7161	78.7161

Fit Plot for x

Data Set	WORK.DIAMOND_LOG
Dependent Variable	price
Selection Method	Forward
Select Criterion	Adj R-Sq
Stop Criterion	Adj R-Sq
Choose Criterion	Adj R-Sq
Effect Hierarchy Enforced	None

Number of Observations Read	53917
Number of Observations Used	53917

Dimensions	
Number of Effects	5
Number of Parameters	5

Forward Selection Summary

Step	Effect Entered	Number Effects In	Adjusted R-Square
* Optimal Value of Criterion			

Forward Selection Summary			
Step	Effect Entered	Number Effects In	Adjusted R-Square
0	Intercept	1	0.0000
1	carat	2	0.8493
2	x	3	0.8537
3	depth	4	0.8567
4	table	5	0.8596*
* Optimal Value of Criterion			

Selection stopped because all effects are in the final model.

Coefficient Progression for price

Fit Criteria for price
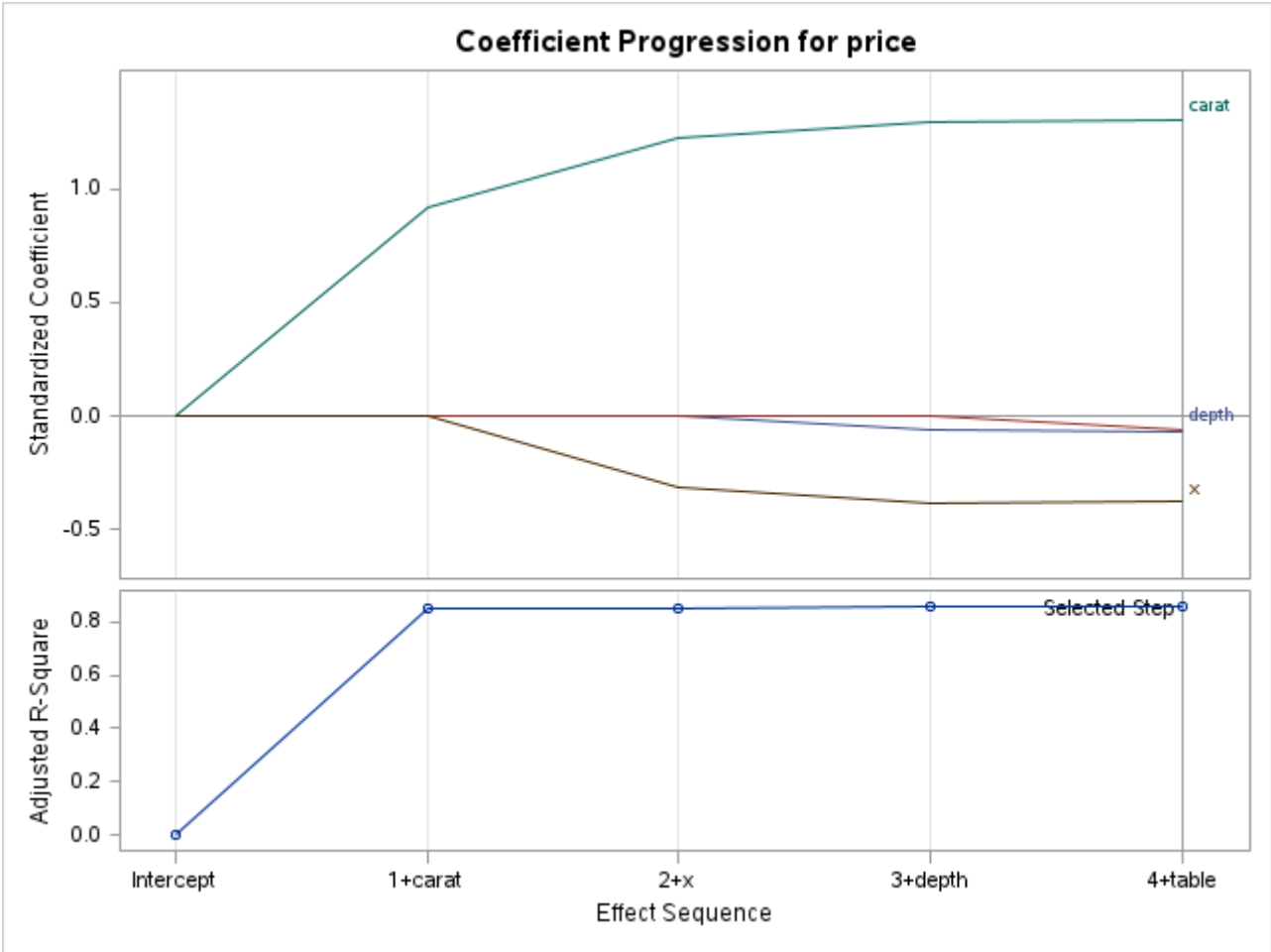
Selected Model

The selected model, based on Adj R-Sq, is the model at Step 4.

Effects:	Intercept depth table carat x

Note: The p-values for parameters and effects are not adjusted for the fact that the terms in the model have been selected and so are generally liberal.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	7.368365E11	1.842091E11	82543.2	<.0001
Error	53912	1.203138E11	2231669		
Corrected Total	53916	8.571503E11			

Root MSE	1493.87726
Dependent Mean	3930.91047
R-Square	0.8596
Adj R-Sq	0.8596
AIC	842097
AICC	842097
SBC	788222

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	21556	422.390770	51.03	<.0001
depth	1	-206.395369	4.857514	-42.49	<.0001
table	1	-102.310943	3.076475	-33.26	<.0001
carat	1	10994	66.953775	164.20	<.0001

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
x	1	-1355.869523	28.326863	-47.87	<.0001

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
x	1	-1355.869523	28.326863	-47.87	<.0001