

## Q1. Loading in new library

<b>Data Set Name</b>	S40840.DIAMONDS	<b>Observations</b>	53940
<b>Member Type</b>	DATA	<b>Variables</b>	11
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	23/07/2021 22:53:22	<b>Observation Length</b>	80
<b>Last Modified</b>	23/07/2021 22:53:22	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	34
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	1635
<b>Obs in First Data Page</b>	1595
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/home/u58705050/sasuser.v94/Assignment 1/diamonds.sas7bdat
<b>Release Created</b>	9.0401M6
<b>Host Created</b>	Linux
<b>Inode Number</b>	228328046
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	u58705050
<b>File Size</b>	4MB
<b>File Size (bytes)</b>	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.

## Q1. Loading in new library

<b>Data Set Name</b>	S40840.DIAMONDS	<b>Observations</b>	53940
<b>Member Type</b>	DATA	<b>Variables</b>	10
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	23/07/2021 22:53:22	<b>Observation Length</b>	80
<b>Last Modified</b>	23/07/2021 22:53:22	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			

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	1596
Number of Data Set Repairs	0
Filename	/home/u58705050/sasuser.v94/Assignment 1/diamonds.sas7bdat
Release Created	9.0401M6
Host Created	Linux
Inode Number	228328047
Access Permission	rw-r--r--
Owner Name	u58705050
File Size	4MB
File Size (bytes)	4587520

Variables in Creation Order					
#	Variable	Type	Len	Format	Informat
1	carat	Num	8	BEST12.	BEST32.
2	cut	Char	11	\$11.	\$11.
3	color	Char	3	\$3.	\$3.
4	clarity	Char	6	\$6.	\$6.
5	depth	Num	8	BEST12.	BEST32.
6	table	Num	8	BEST12.	BEST32.
7	price	Num	8	BEST12.	BEST32.
8	x	Num	8	BEST12.	BEST32.
9	y	Num	8	BEST12.	BEST32.
10	z	Num	8	BEST12.	BEST32.

Q2. Printing Missing Values

Cut Grade	
cut	Frequency
Not Missing	53940

Color Grade	
color	Frequency
Not Missing	53940

Clarity Grade	
clarity	Frequency
Not Missing	53940

Carat of Diamond	
carat	Frequency
Not Missing	53940

Z depth/Z100
--------------

depth	Frequency
-------	-----------

depth	Frequency
Not Missing	53940

table	Frequency
Not Missing	53940

price	Frequency
Not Missing	53940

x	Frequency
Not Missing	53940

y	Frequency
Not Missing	53940

z	Frequency
Not Missing	53940

Q2. As we can see here there appears to be no values missing.

### Q3. Printing First Five Rows.

Obs	carat	cut	color	clarity	depth	table	price	x	y	z
1	0.23	I	Col	SI2	61.5	55	326	3.95	3.98	2.43
2	0.21	P	Col	SI1	59.8	61	326	3.89	3.84	2.31
3	0.23	G	Col	VS1	56.9	65	327	4.05	4.07	2.31
4	0.29	P	Nea	VS2	62.4	58	334	4.2	4.23	2.63
5	0.31	G	Nea	SI2	63.3	58	335	4.34	4.35	2.75

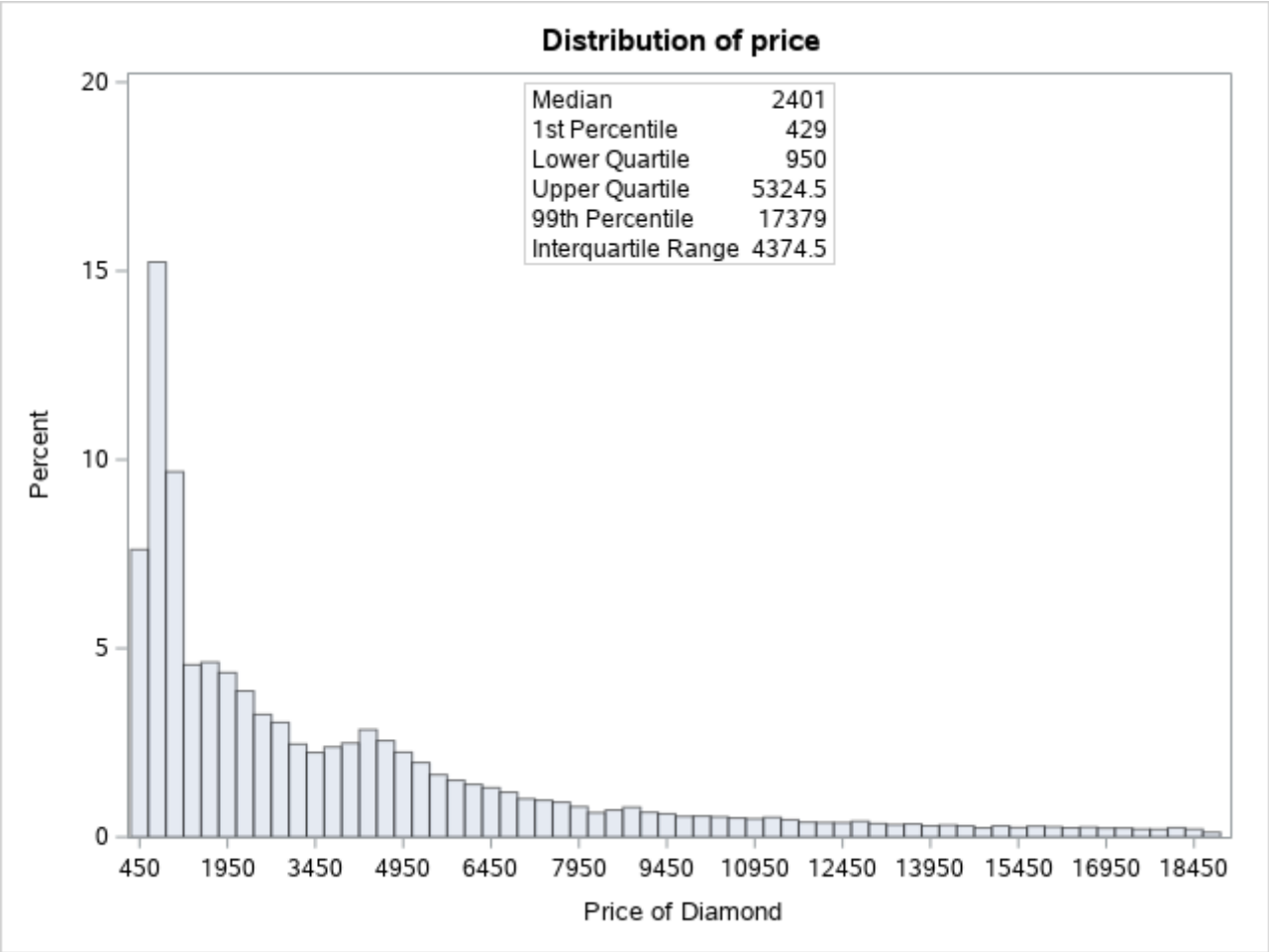
Q3. As we can see this has worked and the colour of the diamond in the 4th row is Nearly Colourless.

### Q5. Printing MEANS and UNIVARIATE procedures.

Analysis Variable : price Price of Diamond					
Median	1st Pctl	Lower Quartile	Upper Quartile	99th Pctl	Quartile Range
2401.00	429.0000000	950.0000000	5324.50	17379.00	4374.50

Q3. As we can see this has worked and the colour of the diamond in the 4th row is Nearly Colourless.

### Q5. Printing MEANS and UNIVARIATE procedures.

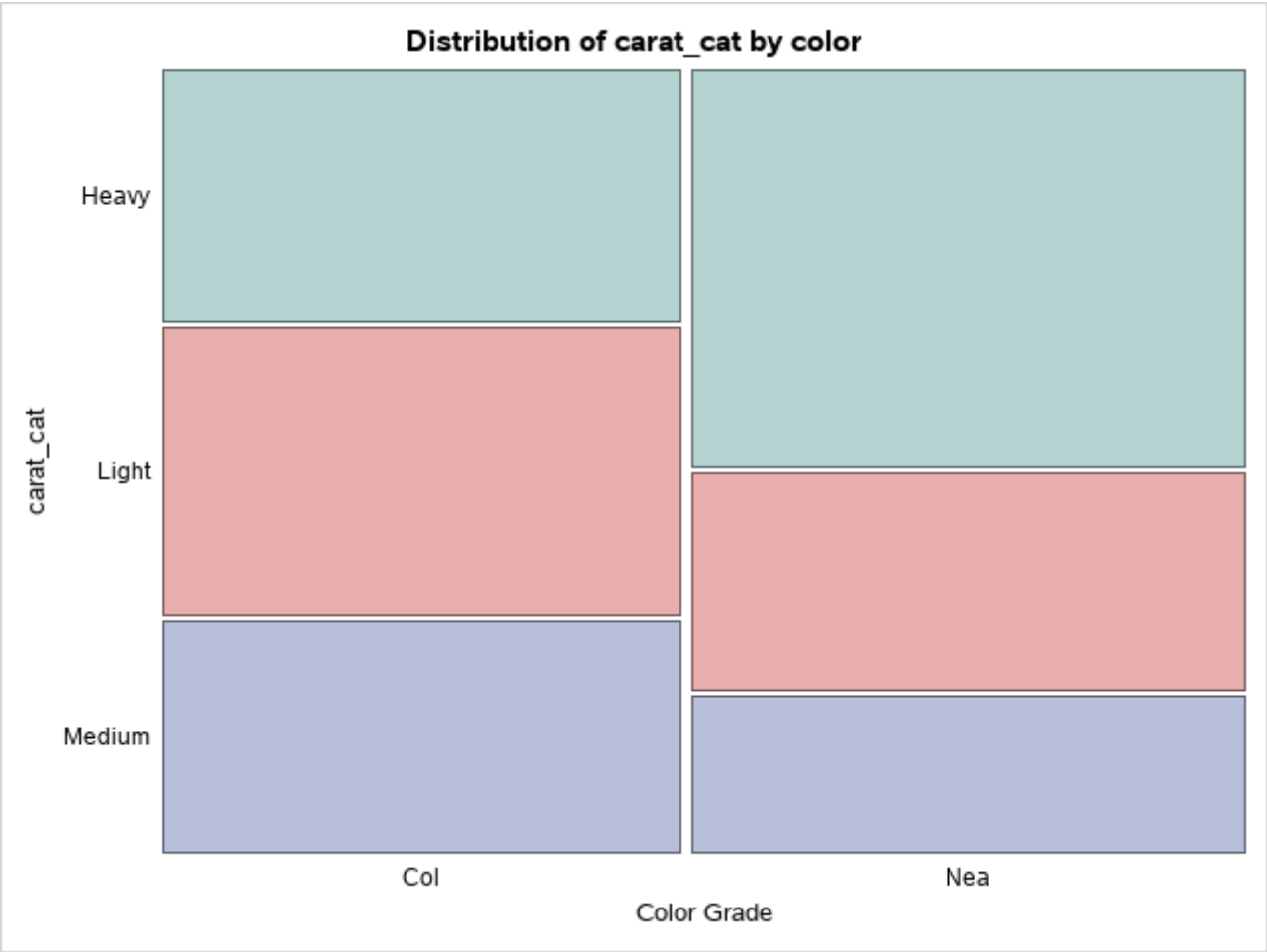
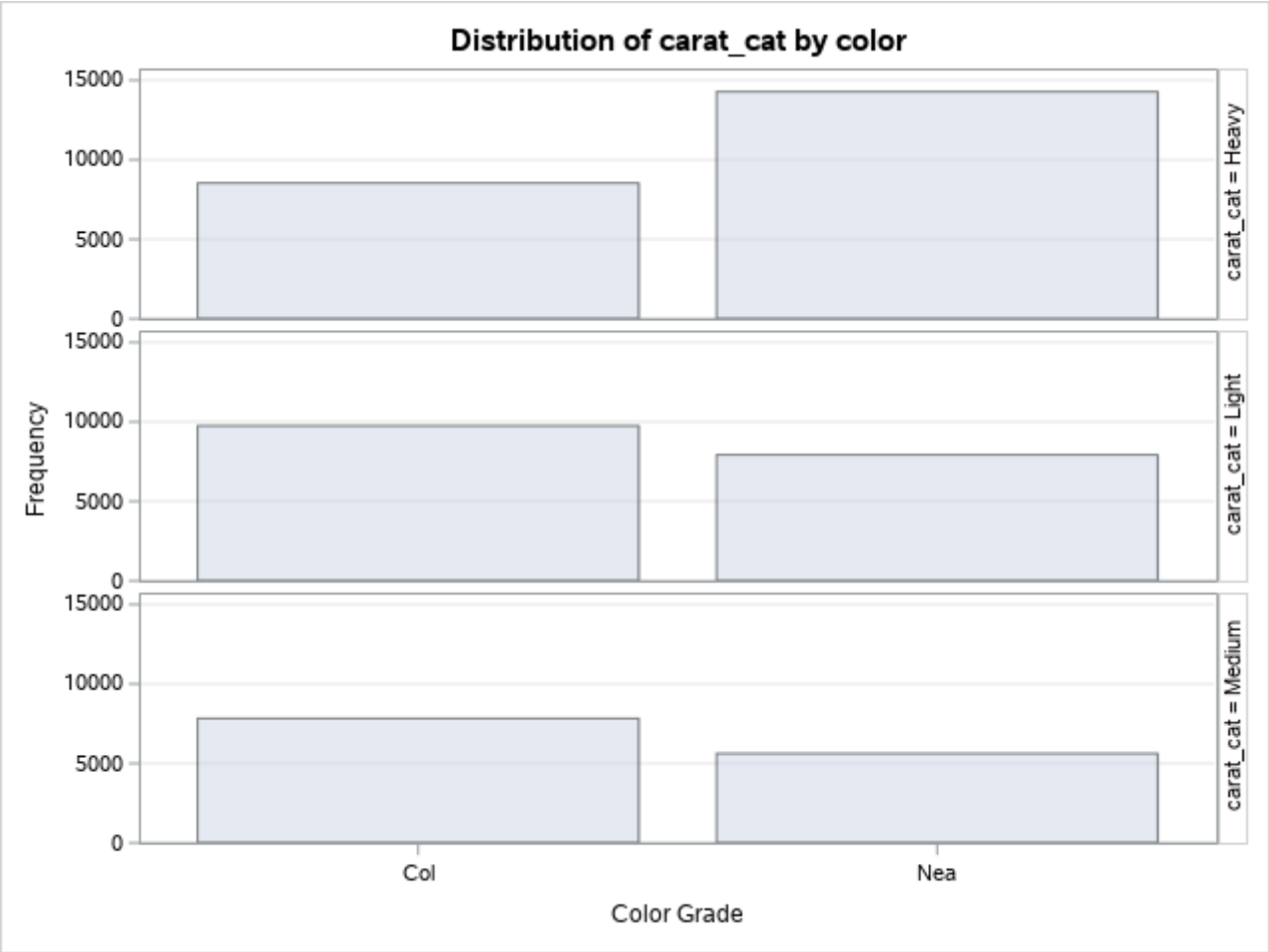


Q5. Here we find right skewed data, which is to be expected from more and more expensive diamonds being in lesser quantity. We should note a large interquartile range which would be indicative of high variance between the prices of diamonds. Furthermore considering our median of 2401, values of over 17000 would indicate some extreme outliers on the far end of prices, especially given we have only specified the 99th percentile.

Q8. Filtering summary statistics

Frequency  
Row Pct  
Col Pct

Table of carat_cat by color			
carat_cat	color(Color Grade)		
	Col	Nea	Total
Heavy	8533	14280	22813
	37.40	62.60	
	32.68	51.32	
Light	9749	7925	17674
	55.16	44.84	
	37.33	28.48	
Medium	7832	5621	13453
	58.22	41.78	
	29.99	20.20	
Total	26114	27826	53940



Q8. The Mean Carat for colourless diamonds of clarity SI2 is 1.9239623.(I am unsure why this prints over Q7, I do not see any problem with my code here)

**Q8. Filtering summary statistics**

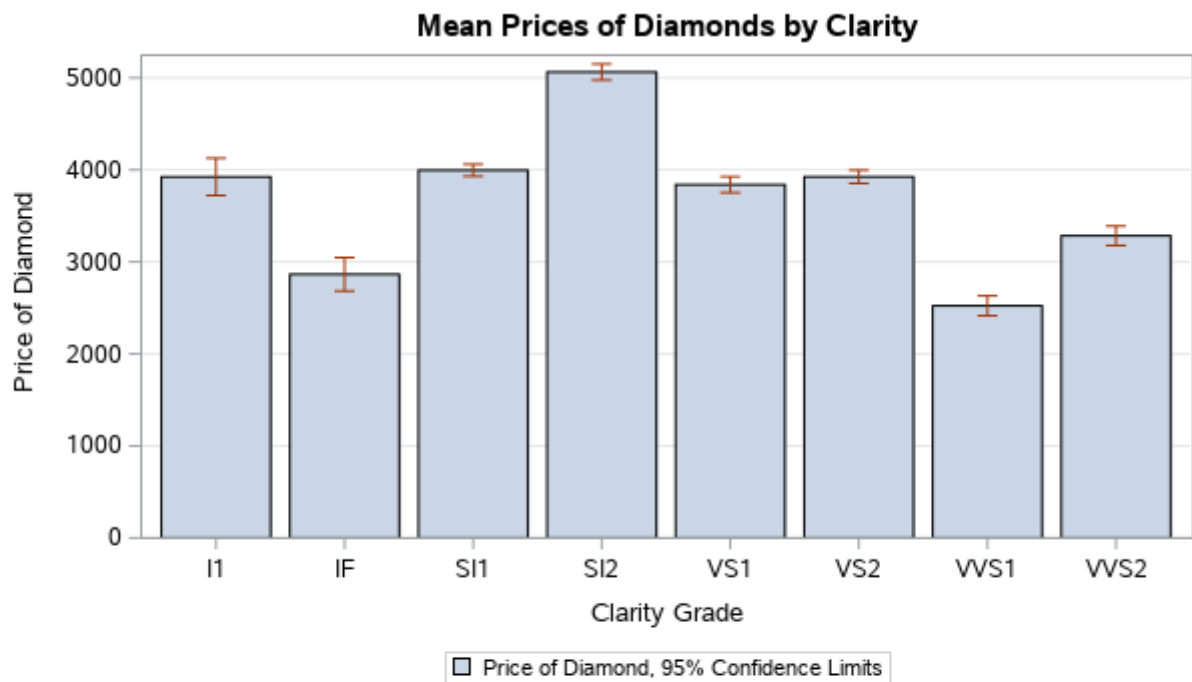
Analysis Variable : carat Carat of Diamond							
Color Grade	Clarity Grade	N Obs	Mean	Std Dev	Minimum	Maximum	N
Col	I1	7	2.6257143	0.5652096	2.0000000	3.4000000	7
	IF	102	1.0941176	0.1093125	1.0000000	1.5100000	102
	SI1	411	1.6083942	0.1545847	1.4400000	2.2100000	411
	SI2	371	1.9239623	0.2456952	1.0100000	2.7100000	371
	VS1	196	1.4087245	0.1608787	1.0300000	1.9300000	196
	VS2	338	1.5140237	0.1146188	1.1000000	2.0000000	338
	VVS1	142	1.1398592	0.1157522	1.0000000	1.5100000	142
	VVS2	231	1.1788312	0.1370335	1.0000000	1.5500000	231
	I1	23	3.3013043	0.7152196	2.3400000	5.0100000	23
	IF	66	1.4474242	0.2488949	1.0900000	2.2900000	66
Nea	SI1	773	1.9180983	0.2570229	1.4900000	2.5700000	773
	SI2	868	2.1421774	0.2216520	1.5000000	3.0400000	868
	VS1	551	1.6691652	0.2471070	1.2100000	2.5900000	551
	VS2	817	1.7767931	0.2693470	1.0100000	3.5100000	817
	VVS1	105	1.4388571	0.2208389	1.1400000	2.3100000	105
	VVS2	221	1.4898190	0.1974818	1.1200000	2.0700000	221

Q8. The Mean Carat for colourless diamonds of clarity SI2 is 1.9239623.(I am unsure why this prints over Q7, I do not see any problem with my code here)

**Q8. Mean price by Clarity**

Analysis Variable : price Price of Diamond		
Clarity Grade	N Obs	Mean
I1	741	3924.17
IF	1790	2864.84
SI1	13065	3996.00
SI2	9194	5063.03
VS1	8171	3839.46
VS2	12258	3924.99
VVS1	3655	2523.11
VVS2	5066	3283.74

Q8. The mean price for SI2 clarity grade diamonds is 5063.03.



Q9. My goal is to create a barchart comparing grades of clarity by price. This will be done using the task and utilities tab and the barchart entry. Here we can see the SI2 grade commands the highest average price whilst grade VVS1 commands the lowest. Price average for each grade seems relatively similar around 4000. Lastly our 95% confidence limits indicate that grade I1 has the highest variance in price as it has the widest limits. This is likely due to I1 being given the smallest sample in our data, as can be observed in the table in Question 5.