THE PROFESSOR PROPOSES

MCDA 5520

Name	Student A# Number
Aman Sharma	A00429273
Bhavya Ahuja	A00428499
Maninder Kaur	A00429943
Santhamohan Manivannan	A00429112
Simon Al Achkar	A00424630
Zewei Yan	A00429842

Table of Content

S. No.	Title	Page No.
1.	Introduction	1
2.	Problem Statement	2
3.	Methodology	3
4.	Regression	4
5.	Variable Entered/Removed	5
6.	Descriptive Statistics	6
7.	Correlations	7
8.	Model Summary	8
9.	Coefficients	9, 10
10.	Anova Table	11
11.	Coefficient Correlation	12
12.	Histogram	13
13.	Relationship Price & Carat	14
14.	Limitations	15
15.	Conclusion	16

INTRODUCTION

Buying a diamond may look simple but it is a complicated and tricky process. This is because of the many classification every diamond has before its price is determined. There are 4 main characteristics known as the 4 C's, namely:

1. Color.
2. Cut.
3. Clarity.
4. Carat.

In addition to these four characteristics, there are further three characteristics which are deemed important; polish, symmetry and certification. Symmetry is the alignment of diamond from it top head to pointy bottom, the better the more expensive. If a diamond is well polished it would be very reflective because a diamond with better polish means it would be more expensive diamond. Certification is another factor, but it is more important for authenticating that whether it's a real diamond or not.

As far, these main characteristics are concerned:

- A diamond can have a color between two shades, yellow or colorless.
- A more perfectly cut diamond would have the right proportions in the facets as well as from all sides. It will also be more expensive.
- Clarity simply measure any detectable defects in the diamonds. If a diamond is flawless it would be more expensive.
- Carat is simply a measurement of diamond's weight. 1 carat of diamond is equal to 0.2 grams. It is highly rare for any diamond to be above 1.5 carats, so if a diamond has more carats it would be more expensive.

PROBLEM STATEMENT

The professor has decided to propose his girlfriend for marriage and therefore he has decided to buy a diamond for his wife-to-be. He thought that he will go to the jewelers shop and will purchase a diamond by keeping the preferences of his girlfriend's taste in his mind. The diamond should be within a range of 2000 to 4000 Canadian dollar according to his budget but the problem he faced when he arrived at the diamond shop are highlighted below:

- He realized that there are various categories of diamonds which are reflected in the price of the diamond.
- Higher the karat of diamond makes the price of diamond higher i.e. big diamonds are rare which increased the value of the diamond.
- The quality and price of diamond depends upon cut, clarity, polish, symmetry and color.
- Different colored diamond are rare to find which increases the value of colorful diamond and made them expensive. Although colorless diamonds are placed with discounts if there is a presence of small yellowish spot.
- These all qualities of diamond confused the professor which made him to think on whether he should buy the diamond at a given price or should conduct regression analysis to find the right price for his selected diamond by collecting data regarding the defined quality and price.

METHODOLOGY

DATA:

The data used in this study is majorly qualitative data, except price and carats, since all other variables measure the quality of the diamond based on different characteristics.

The total number of observation is 440 with the data being collected from 3 different vendors. The dataset, collected by the professor is provided with the case study.

VARIABLES:

Price is taken as a dependent variable, since price is derived from different characteristics of the diamond. Cut, color, clarity and carat are taken as the independent variable.

For carats the data was numerical but for rest of the independent variables data was qualitative. So scaling of the qualitative dataset was completed on a statistical software for the purpose of establishing a model.

Another requirement was to use color and clarity as dummy variables, which only take the value of 0 or 1. We also used cut as a dummy variable for developing the model. The conditions were set according to the professor's preference which are as follows:

i. Clarity will take the value 1 when its scale is SI2, otherwise for all other conditions it will take a value of 0.

ii. Color will take the value 1 when its scale is "J", otherwise for all other conditions it will take a value of 0.

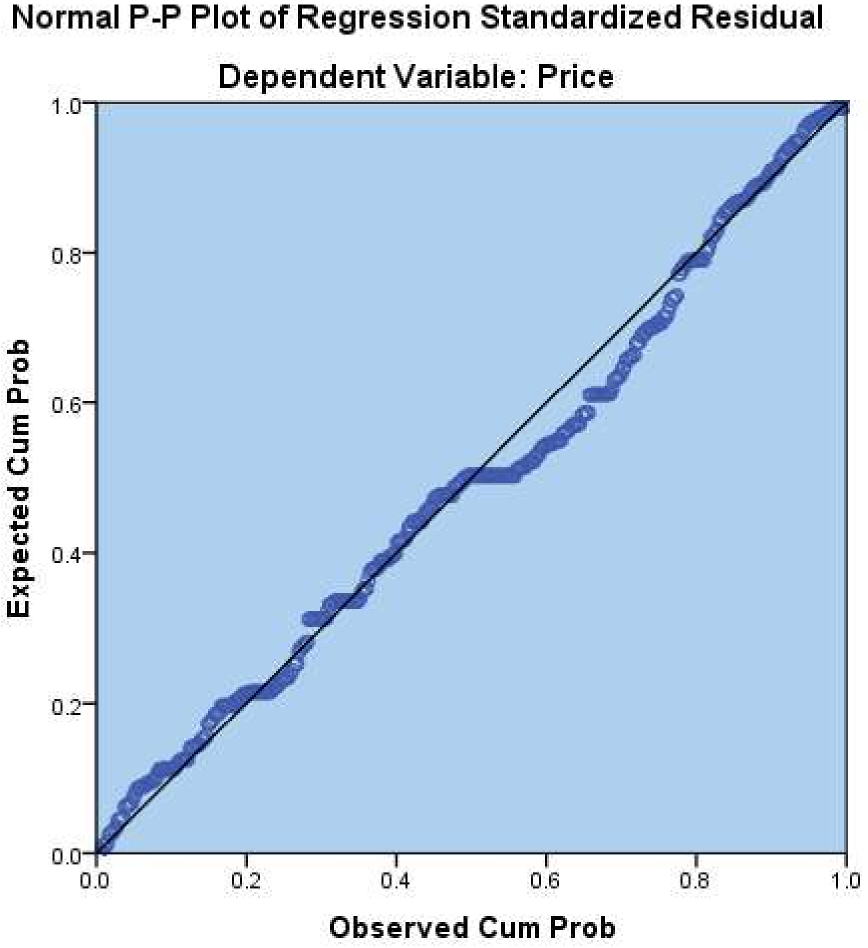
iii. Cut will take the value 1 when its scale is "very good", otherwise for all other conditions it will take a value of 0.

REGRESSION

Linear regression was the method which was used to regress the independent variables against the dependent variable. This method was considered after plotting the data, we can easily see that a straight line fits through the data with minimum squared errors. So, a linear regression model was the best choice.

INTERPRETATION OF DATA REGARDING REGRESSION MODEL:

Following is the data in analytical form to build regression model for the professor to determine price of the diamond regarding his preferred quality characteristics.



VARIABLES ENTERED/REMOVED

Price is taken as a dependent variable. Carats, Clarity, Color and Cut are used as an independent variable, while the qualitative data for clarity, color and cut have been scaled and then transformed into dummy variables i.e. they take values of 1 and 0.

- More specifically, clarity will take the value 1 when its scale is SI2,
- color will take the value 1 when its scale is “J” and
- cut will take the value 1 when its scale is “very good” otherwise, all values will be zero.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Clarity Dummy, Color Dummy, Cut Dummy, Carat[b]	.	Enter

a. Dependent Variable: Price

b. All requested variables entered.

DESCRIPTIVE STATISTICS

The number of observation for each variable is 440. The data used for analysis consist of all three vendors/wholesalers combined. The descriptive statistics are as follow,

- The mean value for Price is 1716.74 with a standard deviation of 1175.689. One possible reason for such a high standard deviation can be that the three different vendors might have a high difference of prices among themselves.
- Carat has a mean value of 0.6692 with a standard deviation of 0.37980. This is consistent with the fact that high carat diamonds are very rare.
- Cut has a mean of 0.22 with a standard deviation of 0.415.
- Color has a mean value of 0.23 with a standard deviation of 0.424.
- Clarity has a mean value of 0.25 with a standard deviation of 0.434.

Descriptive Statistics

	Mean	**Std. Deviation**	**N**
Price	1716.74	1175.689	440
Carat	.6692	.37980	440
Color_Dummy	.23	.424	440
Clarity_Dummy	.25	.434	440
Cut_Dummy	.22	.415	440

CORRELATIONS

Pearson Correlation is a measure of positive or negative correlation between two variables, without taking into account dependent and independent variable. This test usually take the values between -1 and 1. This test is useful in selecting the most relevant variables and it is a normal to exclude variables with high correlation i.e. more than 0.7.

Considering this test for our data I see that none of the dependent variable are highly correlated with each other, with number of observations being 440. Therefore, all variable can be included in our analysis without any doubt. The interpretation for this table is as follow,

- Carat and cut are negatively correlated with a value of -0.328. This means the higher the carat the more difficult it would be to cut it.
- Carat and color are positively correlated with a value of 0.175.
- Carat and clarity are positively correlated with a value of 0.012.
- Cut and color are negatively correlated with a value of -0.035.
- Cut and clarity are positively correlated with a value of 0.035.
- Color and clarity are negatively correlated with a value of -0.009.

Correlations

		Price	Carat	Color_Dummy	Clarity_Dummy	Cut_Dummy
Pearson Correlation	Price	1.000	.925	.112	.127	-.244
	Carat	.925	1.000	.175	.012	-.328
	Color_Dummy	.112	.175	1.000	-.009	-.035
	Clarity_Dummy	.127	.012	-.009	1.000	.035
	Cut_Dummy	-.244	-.328	-.035	.035	1.000
Sig. (1-tailed)	Price	.	.000	.009	.004	.000
	Carat	.000	.	.000	.398	.000
	Color_Dummy	.009	.000	.	.423	.232
	Clarity_Dummy	.004	.398	.423	.	.233
	Cut_Dummy	.000	.000	.232	.233	.
N	Price	440	440	440	440	440
	Carat	440	440	440	440	440
	Color_Dummy	440	440	440	440	440
	Clarity_Dummy	440	440	440	440	440
	Cut_Dummy	440	440	440	440	440

MODEL SUMMARY

Our model is a multiple regression analysis, therefore it will be more efficient and effective to interpret R2 instead of R since multiple variables are involved in our analysis. Also, our analysis of Peterson correlation has proved that our variables are significant enough to be included in our analysis.

Our coefficient of determination is 0.875 or 87.5%. Coefficient of determination determines the percentage variation in the dependent variable which is explained by all the dependent variables together. This is another sign that our regression model is strong enough to predict the prices of diamonds since our model explains 87.5% of variation in the prices of diamonds.

Since our data is not a time series data, no serial correlation can be expected. Hence, the Durbin-Watson would be irrelevant.

Model Summary[b]

					Change Statistics					
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**	**R Square Change**	**F Change**	**df1**	**df2**	**Sig. F Change**	**Durbin-Watson**
1	.936[a]	.876	.875	416.492	.876	765.784	4	435	.000	.474

a. Predictors: (Constant), Cut_Dummy, Clarity_Dummy, Color_Dummy, Carat

b. Dependent Variable: Price

COEFFICIENTS

Moving forward with the interpretations of coefficients, they are interpreted as follow,

- The constant (y-intercept) is -341.343, which means that when all the variables are 0 the price of diamond would be -341.343. This is not a realistic measure since prices cannot be negative, but this is just one extreme. So this negative number would be ignored and even if the values of all independent variables is 0 the price would also be 0.
- The coefficient on carat is 2952.297, this means that when there is a one unit increase in carat the prices would change (increase) by 2952.297. There is a standard deviation of 56.269. The confidence interval at 95% is 2841.704 – 3062.89.
- The coefficient on cut is 179.073, this means that when there is a one unit increase in carat the prices would change (increase) by 179.073. There is a standard deviation of 50.759. The confidence interval at 95% is 79.309 – 278.837.
- The coefficient on color is -142.004, this means that when there is a one unit increase in carat the prices would change (decrease) by -142.004. There is a standard deviation of 47.641. The confidence interval at 95% is -235.639 – -48.370.
- The coefficient on clarity is 304.083, this means that when there is a one unit increase in carat the prices would change (decrease) by 304.083. There is a standard deviation of 45.900. The confidence interval at 95% is 213.869 – 394.297.

In accordance with the obtained results the regression model created is as follow,

Y = -341.343 + 2952.297X + 179.073X2 - 142.004X3 + 304.083X4

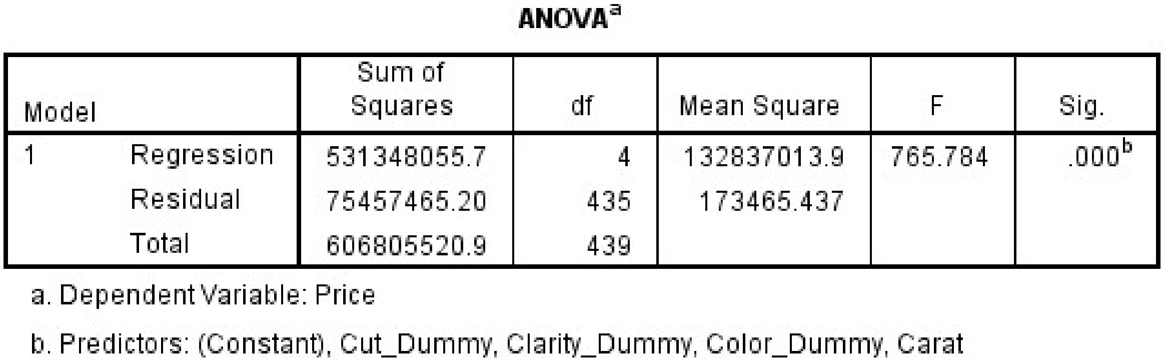
Coefficients[a]

	Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
Model	B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1 (Constant)	-341.343	47.831		-7.136	.000	-435.352	-247.333
Carat	2952.297	56.269	.954	52.467	.000	2841.704	3062.890
Cut_Dummy	179.073	50.759	.063	3.528	.000	79.309	278.837
Color_Dummy	-142.004	47.641	-.051	-2.981	.003	-235.639	-48.370
Clarity_Dummy	304.083	45.900	.112	6.625	.000	213.869	394.297

a. Dependent Variable: Price

ANOVA TABLE

The p-value of the ANOVA table indicates that significance level of 5% is greater than the p-value and therefore we conclude that our tested results are significant and rely upon the conditions that the regression model will determine different results for different qualities of the diamond. High value of statistics also shows that our significant level is high and independent variables are taken to analyze the regression coefficient explained the greater part of the regression equation and residual term is low which indicates unexplained part or error in explaining the price through regression model is low which makes our regression model significant.



ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	531348055.7	4	132837013.9	765.784	.000[b]
	Residual	75457465.20	435	173465.437		
	Total	606805520.9	439			

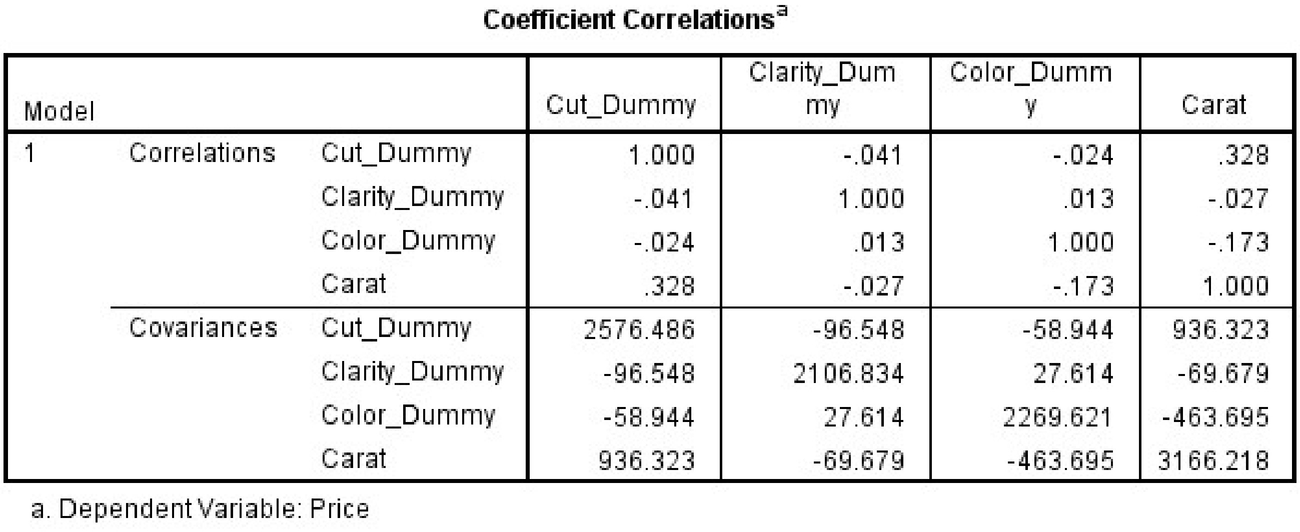
a. Dependent Variable: Price

b. Predictors: (Constant), Cut_Dummy, Clarity_Dummy, Color_Dummy, Carat

COEFFICIENT CORRELATION

Correlation is a standardized form of covariance which determines that how the value of coefficient reacts when the value of other coefficient changes. In this scenario the correlation coefficient measures the relationship between characteristics of diamonds with each other. These coefficient correlations are highlighted below.

- Clarity dummy is slightly positively correlated with color dummy and slightly negative correlated with cut dummy and carat dummy. This correlation indicates that the value of clarity dummy increases with the color dummy and decreases with cut dummy and carat.
- The correlation coefficient indicates that the value of color dummy decreases with the increase in cut dummy and carat.
- The value of cut dummy increases with the increase in carat but decreases with the increase in clarity dummy and color dummy.
- The value of carat decreases with the clarity dummy and color dummy and increases with the increase in cut dummy.



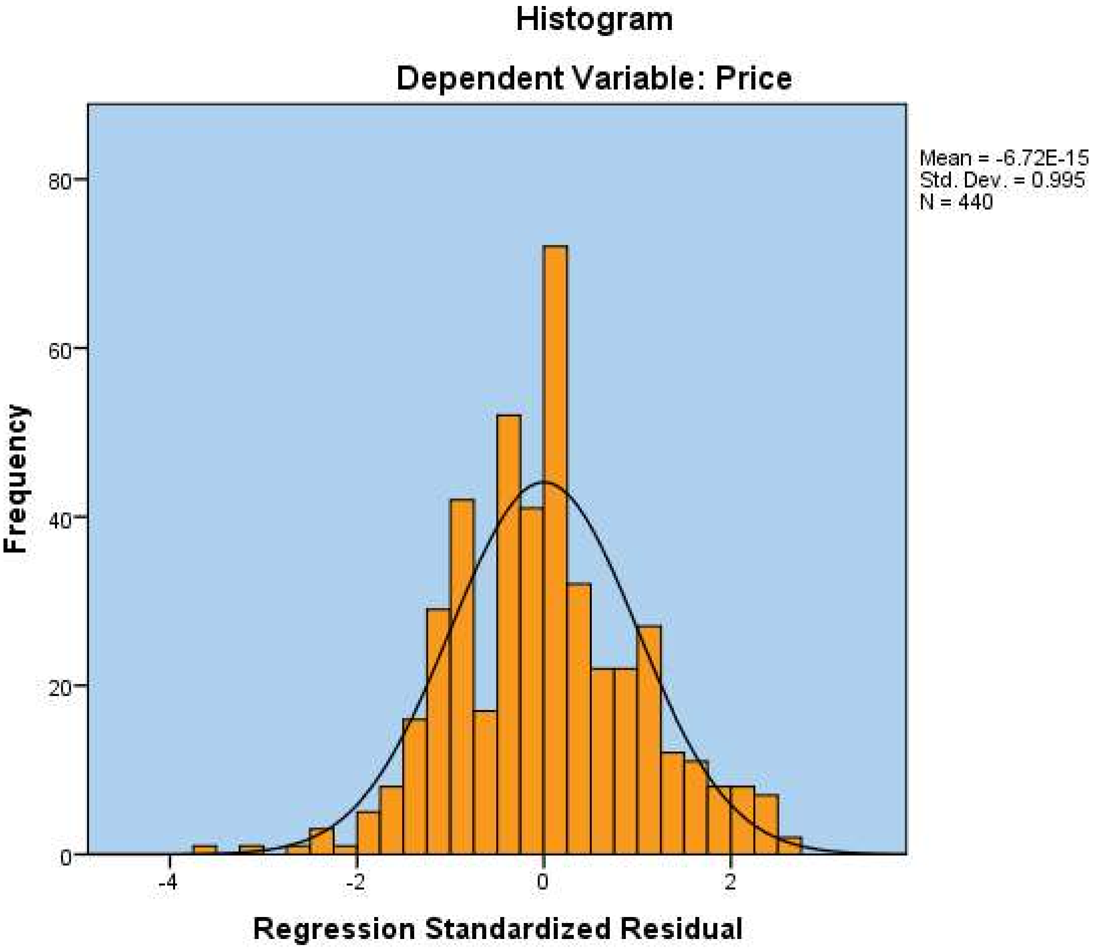
Coefficient Correlations[a]

Model			Cut_Dummy	Clarity_Dummy	Color_Dummy	Carat
1	Correlations	Cut_Dummy	1.000	-.041	-.024	.328
		Clarity_Dummy	-.041	1.000	.013	-.027
		Color_Dummy	-.024	.013	1.000	-.173
		Carat	.328	-.027	-.173	1.000
	Covariances	Cut_Dummy	2576.486	-96.548	-58.944	936.323
		Clarity_Dummy	-96.548	2106.834	27.614	-69.679
		Color_Dummy	-58.944	27.614	2269.621	-463.695
		Carat	936.323	-69.679	-463.695	3166.218

a. Dependent Variable: Price

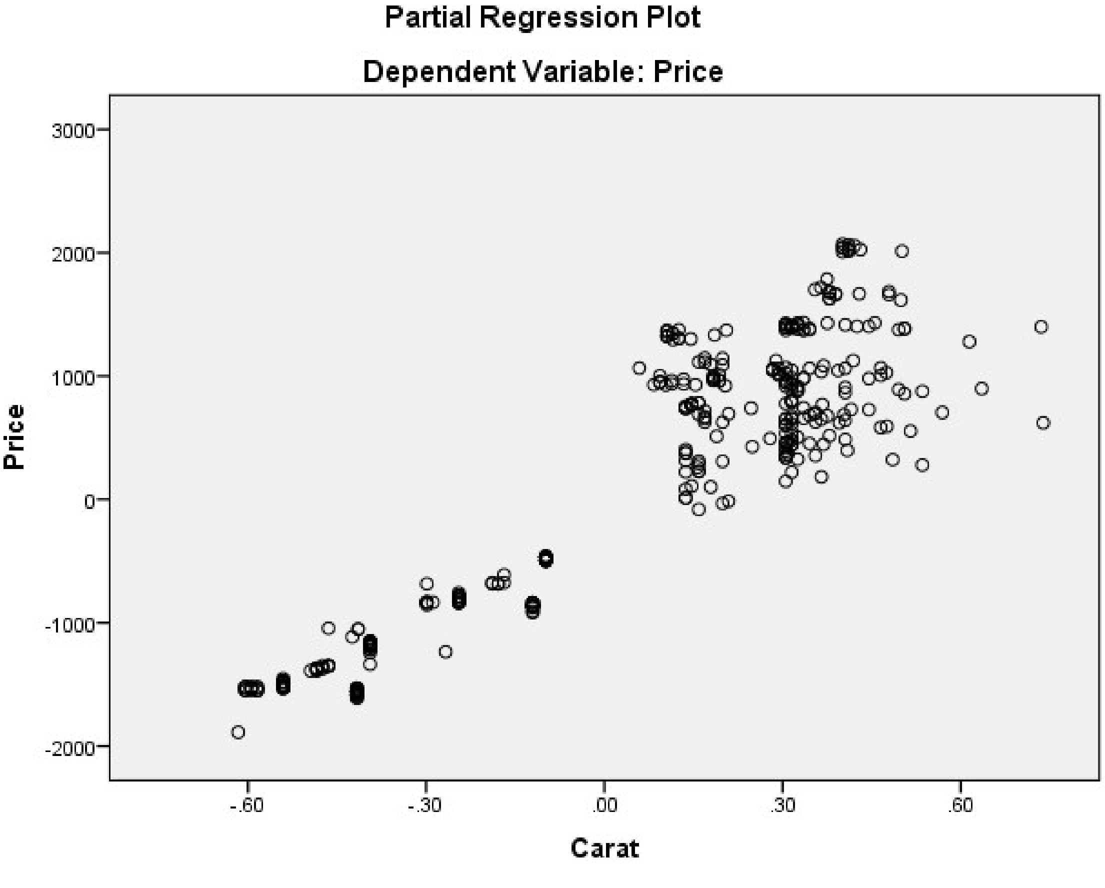
HISTOGRAM

The histogram bar chart shows that the data is distributed symmetrically which means that it is distributed equally. The data is skewed in the right direction indicating that most of the data is positive. The data also shows that there are outliers presented in the data which indicates these outliers are affecting the mean of the data which could generate ambiguity in analyzing the data.



RELATIONSHIP BETWEEN PRICE AND CARAT

The below partial regression scatter plot diagram for the regression equation formulated indicates that when the carats of the diamond increases, the price of the diamonds also increases i.e. the more the carat the higher the price of the diamond.



LIMITATIONS

One of the limitations of our model is that I have used carat, clarity and cut as dummy variables with the conditions that,

- Clarity will take the value 1 when its scale is SI2, otherwise for all other conditions it will take a value of 0.
- Color will take the value 1 when its scale is "J", otherwise for all other conditions it will take a value of 0.
- Cut will take the value 1 when its scale is "very good", otherwise for all other conditions it will take a value of 0.

The model has been conditioned this way because of the professor's preference of the diamond. This would give us the correct price for the specified diamond, according to the above regression model.

CONCLUSION

The professor was looking for a diamond with the following specifications:

- Carat Weight: 0.9.
- Cut: Very Good.
- Color: J (Slightly Yellow).
- Clarity: SI2 (Slightly included: very few inclusions at 10x)
- Polish: Good.
- Symmetry: Very Good.
- Certification: GIA (Gemological Institute of America)

He was quoted a price of $3,100. When the above regression model is used with the same specifications in the model i.e.

Y = -341.343 + 2952.297X + 179.073X2 - 142.004X3 + 304.083X4,

the following results are obtained,

Constant	X	X_2	X_3	X_4
-341.343	2952.297	179.073	-142.004	304.083
X_1 =	0.9			
X_2 =	1			
X_3 =	1			
X_4 =	1			
Price =	**2656.876**			

Therefore, according to the model the professor should get the diamond for about $2,656.87.