STATISTICS & ITS APPLICATION IN BUSINESS

MCDA-5530

Master of Science in Computing and Data Analytics

Final Project Report

The Professor Proposes

Submitted by:

Sonam Vadsaria,	A00431604
Sunil Padikar,	A00428089
Jasleen Kour,	A00425843
Sri Akhil Reddy Kovvuri,	A00428260
Sadman Hoque Sadi,	A00426020
Parijat Bandyopadhyay,	A00430847

Submitted to:

Dr. Michael Zhang

SAINT MARY'S UNIVERSITY SINCE 1802

One University. One World. Yours.

Table of Contents

Introduction ***3***

Diamond Characteristics **3**

Analysis ***4***

Carat **4**

Descriptive Statistics: ***5***

Price: **5**

Carat **5**

Color: **6**

Clarity: **6**

Cut: **7**

Certification, Summary and Polish **7**

Feature Engineering ***8***

Color vs Price **8**

Clarity vs Price **10**

Cut vs Price **12**

Polish vs Price **14**

Certification vs Price **16**

Symmetry vs Price **17**

Multicollinearity ***19***

Final Model with All above variables ***20***

Final Model ***20***

Summary of our model ***22***

Conclusion ***22***

Introduction

The study originates from a professor wanting to buy a diamond ring for his to-be wife in the price range of $2000-$4000, however, he was quickly met with a complicated array of possible configurations regarding diamonds which necessitated the use of statistical analysis in order to find a ring which was of the right value for its configuration. The professor consulted three diamond wholesalers in order to get a proper understanding of diamond configurations and the price range and characteristics of the diamonds each wholesaler offered.

Eventually the professor made the decision of buying a ring costing $3100 consisting of the following properties:

- Carat: 0.9
- Cut: Very Good
- Color: J
- Clarity: SI2
- Polish: Good
- Symmetry: Very Good
- Certification: GIA

Diamond Characteristics

Based on the research the professor has done diamonds have 7 key characteristics based on which their quality and subsequent price is decided, of which the first three are the most important as per the wholesalers. They are:

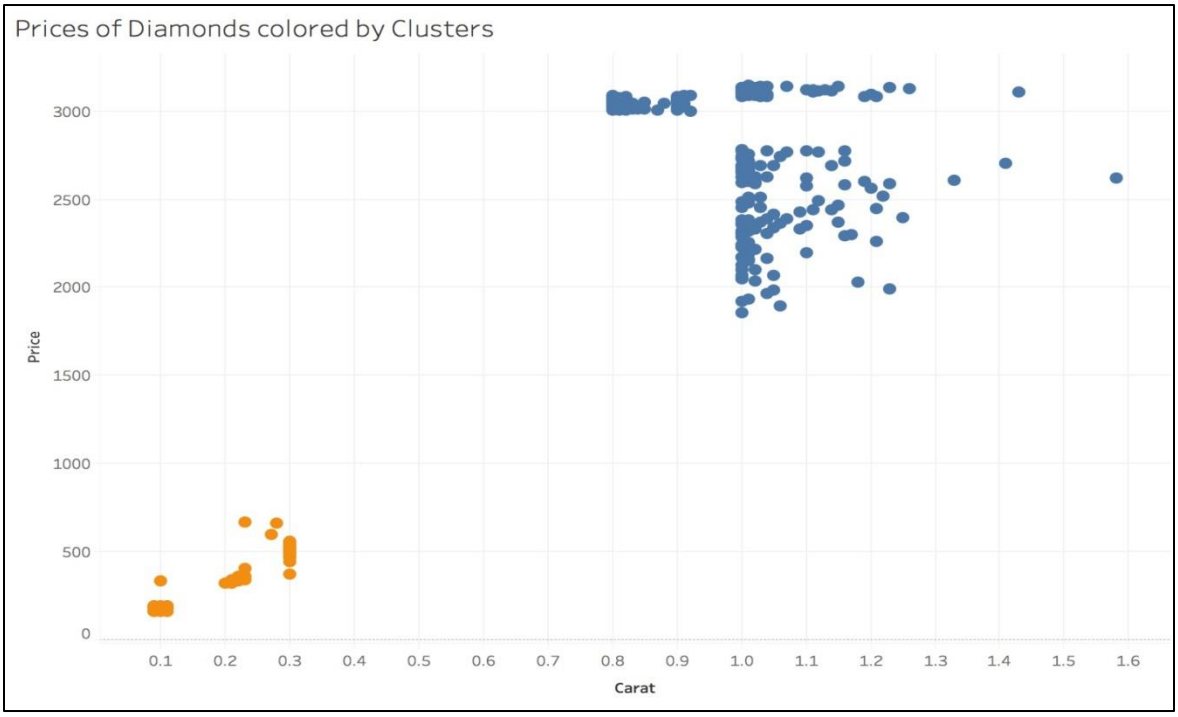
- **Carat:** Not to be confused with Karat, used for gold, this is the weight of the diamond where 1 carat equals 0.2g. Generally having a linear relation with price, however the ratio gets significantly steeper above 1.5carats.
- **Color:** Diamonds can be found in numerous colors but most commonly they are white, yellow and shades in between. This supposedly has a general correlation with price but often seen as a matter of personal choice. In the samples used for the analysis the diamond color was categorized from ‘D’ to ‘Z’ as being completely colorless to being perfectly Yellow in color.
- **Cut:** This refers to the proportions in which the diamond has been cut in and whether the cuts are of the correct depth. The cut is particularly important since this is what gives the reflective properties of the diamond. In the study samples the cut was categorized as from ‘Poor’ to ‘Excellent’.
- **Clarity:** Simply the noticeable flaws on the surface of the diamond and within.
- **Polish and Symmetry:** Like the cut, the polish and symmetry affect the reflective properties of the diamond and thus how ‘brilliant’ it looks to the eye. The measure of each is categorized from ‘Poor’ to ‘Ideal’.
- **Certification:** The certification of the diamond characteristics also has some affect on the price based on the reputation of the organization doing the certification. The most reputable one’s are GIA and AGS.

Analysis

The process of analysis is based on carrying on linear regression analysis on every category of every diamond property listed above. This is because each of the properties and their categories are the factors which affect the final price of the diamond, however they do not do so at the same level. Hence multiple linear regressions were carries out I order to find the significance of each in respect to changes in price. Depending on this analysis clustering has been carries out for each sub category.

Carat

The first analysis done was on the price against carat ratio. For this we have drawn out a scatter plot to get a visual understanding of exactly how the carat is effecting the price.



Based on our initial observation it is rather apparent that the offers of diamond the professor can buy from the three wholesalers consulted fall into 2 apparent clusters, and that it has a very linear effect on the price. The diamond that the professor ultimately bought falls into the blue cluster, and thus for the rest of our analysis we only considered the diamonds which are on that cluster.

Descriptive Statistics:

To understand behavior of data descriptive Analysis was run on given data that has number of observation as 440. Analyzed data included data for all the three wholesalers. Following are the observations made after running descriptive statistics over provided data:

Price:

When considering Price which has numerical data, standard deviation came out to be $1,175.68913 and mean value was $1,716.7386. The maximum and minimum price of a diamond are 3145 and 160 respectively. Results can be seen in the figure below.

Statistics

Price

N	Valid	440
	Missing	0
Mean		$1,716.7386
Median		$2,169.0000
Std. Deviation		$1,175.68913
Range		$2,985.00
Minimum		$160.00
Maximum		$3,145.00

Carat

In case of carrot (numerical data), standard deviation came out to be 0.12264 and mean value was 1.0032. The maximum and minimum Carat of a diamond are 1.58 and 0.8 respectively. Results can be seen in the figure below.

Statistics

Carat

N	Valid	240
	Missing	0
Mean		1.0032
Median		1.0100
Std. Deviation		.12264
Range		.78
Minimum		.80
Maximum		1.58

Color:

In case of color which is categorical data provided after running descriptive statistics we saw that categories H, I and J that are having highest maximum frequency with 71,79 and 72 respectively.

Colour

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	D	20	4.5	4.5	4.5
	E	54	12.3	12.3	16.8
	F	58	13.2	13.2	30.0
	G	43	9.8	9.8	39.8
	H	71	16.1	16.1	55.9
	I	79	18.0	18.0	73.9
	J	72	16.4	16.4	90.2
	K	31	7.0	7.0	97.3
	L	12	2.7	2.7	100.0
	Total	440	100.0	100.0	

Clarity:

In case of clarity descriptive statistics gave the results shown in the figure below. We saw that categories SI1(Slightly Included: very, very few inclusions at 10×) and SI2 (Slightly Included: very few inclusions at 10×) has the highest frequency of 116 and 110 respectively among rest of categories clarity is distributed in.

Clarity

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	I1	82	18.6	18.6	18.6
	I2	28	6.4	6.4	25.0
	SI1	116	26.4	26.4	51.4
	SI2	110	25.0	25.0	76.4
	SI3	26	5.9	5.9	82.3
	VS1	30	6.8	6.8	89.1
	VS2	41	9.3	9.3	98.4
	VVS1	2	.5	.5	98.9
	VVS2	5	1.1	1.1	100.0
	Total	440	100.0	100.0	

Cut:

In case of Cut descriptive statistics gave the results shown in the figure below. We saw that categories I, V and X (i.e. Ideal, Very Good and Excellent respectively) have the highest frequency of 86, 97 and 149 respectively among rest of categories cut is distributed in.

Cut

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	F	59	13.4	13.4	13.4
	G	49	11.1	11.1	24.5
	I	86	19.5	19.5	44.1
	V	97	22.0	22.0	66.1
	X	149	33.9	33.9	100.0
	Total	440	100.0	100.0	

Certification, Summary and Polish

Likewise, if we consider the case of certification, symmetry and polish there the descriptive statistics came with results shown below. Certification EGL (European Gemological Laboratories) and GIA (Gemological Institute of America) have the highest frequency of 119 and 265 respectively. In symmetry categories G (Good) and V (Very Good) have the highest frequency values with 157 and 3206 respectively. At last Polish has the highest frequency values of categories G (Good) and V (Very Good).

Certification

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	AGS	12	2.7	2.7	2.7
	DOW	1	.2	.2	3.0
	EGL	119	27.0	27.0	30.0
	GIA	265	60.2	60.2	90.2
	IGI	43	9.8	9.8	100.0
	Total	440	100.0	100.0	

Symmetry

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	F	21	4.8	4.8	4.8
	G	157	35.7	35.7	40.5
	I	5	1.1	1.1	41.6
	V	206	46.8	46.8	88.4
	X	51	11.6	11.6	100.0
	Total	440	100.0	100.0	

Polish

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	F	5	1.1	1.1	1.1
	G	165	37.5	37.5	38.6
	I	5	1.1	1.1	39.8
	V	204	46.4	46.4	86.1
	X	61	13.9	13.9	100.0
	Total	440	100.0	100.0	

Feature Engineering

As an initial analysis we have done featuring engineering in order to understand how each independent variable (Carat, Symmetry, Polish, Cut, Clarity, Certification, Color) is affecting the price of diamond.

Color vs Price

By running regression model by taking price as dependent variable and color as Independent variable we have the following results as shown in below snap.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.256[a]	.066	.054	$357.36472	1.003

a. Predictors: (Constant), ColourN=near colourless, ColourN=Very light yellow, ColourN=faint yellow

b. Dependent Variable: Price

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2118111.121	3	706037.040	5.528	.001[b]
	Residual	30139452.73	236	127709.546		
	Total	32257563.85	239			

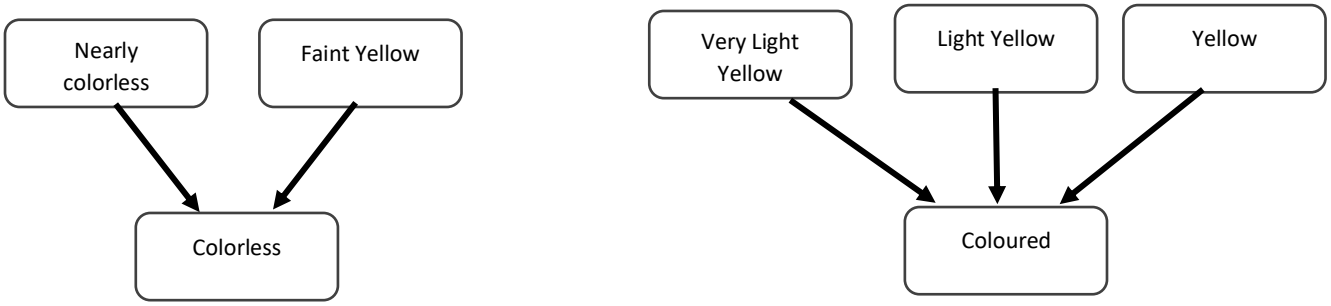
a. Dependent Variable: Price

b. Predictors: (Constant), ColourN=near colourless, ColourN=Very light yellow, ColourN=faint yellow

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	2814.615	44.326		63.499	.000
	ColourN=Very light yellow	-446.115	112.282	-.265	-3.973	.000
	ColourN=faint yellow	-78.959	62.216	-.097	-1.269	.206
	ColourN=near colourless	-32.480	57.403	-.043	-.566	.572

From the above result, color “nearly colourless” and “faint yellow” are highly insignificant in determining price of diamond. As a further analysis colours are grouped into two categories as below.



Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.148[a]	.022	.018	$364.10515	1.007

a. Predictors: (Constant), Colour_2Cat=1.0

b. Dependent Variable: Price

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	705293.813	1	705293.813	5.320	.022[b]
	Residual	31552270.04	238	132572.563		
	Total	32257563.85	239			

a. Dependent Variable: Price

b. Predictors: (Constant), Colour_2Cat=1.0

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2795.248	28.696		97.411	.000
	Colour_2Cat=1.0	-115.362	50.016	-.148	-2.307	.022

a. Dependent Variable: Price

The above result shows model for colors with two categories (Colorless and Color). Now, Color has significant effect on price of diamond.

Clarity vs Price

Here Clarity is independent variable and price is dependent variable. By applying these variables in the regression model we got the results as shown in below snap.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.632[a]	.399	.384	$287.94626	1.663

a. Predictors: (Constant), ClarityN=8.0, ClarityN=7.0, ClarityN=4.0, ClarityN=6.0, ClarityN=2.0, ClarityN=5.0

b. Dependent Variable: Price

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12776406.98	6	2129401.163	25.682	.000[b]
	Residual	19235826.89	232	82913.047		
	Total	32012233.87	238			

a. Dependent Variable: Price

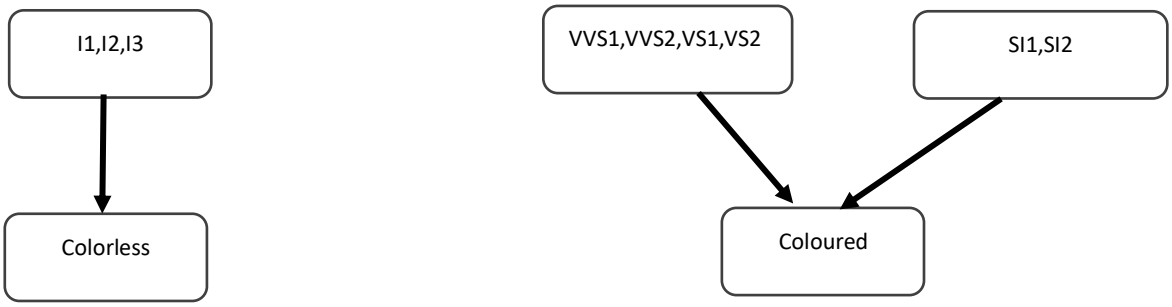
b. Predictors: (Constant), ClarityN=8.0, ClarityN=7.0, ClarityN=4.0, ClarityN=6.0, ClarityN=2.0, ClarityN=5.0

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	2622.911	32.396		80.963	.000
	ClarityN=2.0	-278.060	64.190	-.241	-4.332	.000
	ClarityN=4.0	-3.527	65.104	-.003	-.054	.957
	ClarityN=5.0	368.012	48.219	.447	7.632	.000
	ClarityN=6.0	376.607	64.190	.326	5.867	.000
	ClarityN=7.0	380.517	113.553	.175	3.351	.001
	ClarityN=8.0	431.714	106.835	.212	4.041	.000

a. Dependent Variable: Price

From the above result, clarity “SI3” are highly insignificant in determining price of diamond. As a further analysis clarity is grouped into two categories as below.



Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.381[a]	.145	.143	$1,088.20048	.098

a. Predictors: (Constant), Clarity2cat=1.0

b. Dependent Variable: Price

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	88134555.22	1	88134555.22	74.427	.000[b]
	Residual	518670965.7	438	1184180.287		
	Total	606805520.9	439			

a. Dependent Variable: Price

b. Predictors: (Constant), Clarity2cat=1.0

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	1458.342	59.903		24.345	.000
	Clarity2cat=1.0	1033.585	119.807	.381	8.627	.000

a. Dependent Variable: Price

The above result shows model for clarity with two categories (Visible and 30x-10x). Now, Clarity has significant effect on price of diamond.

Cut vs Price

Here Cut is independent variable and price is dependent variable. By applying these variables in the regression model we got the results as shown in below snap.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.379[a]	.144	.129	$342.78924	1.548

a. Predictors: (Constant), CutN=Ideal, CutN=Very Good, CutN=good, CutN=Excellent

b. Dependent Variable: Price

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4644014.751	4	1161003.688	9.881	.000[b]
	Residual	27613549.10	235	117504.464		
	Total	32257563.85	239			

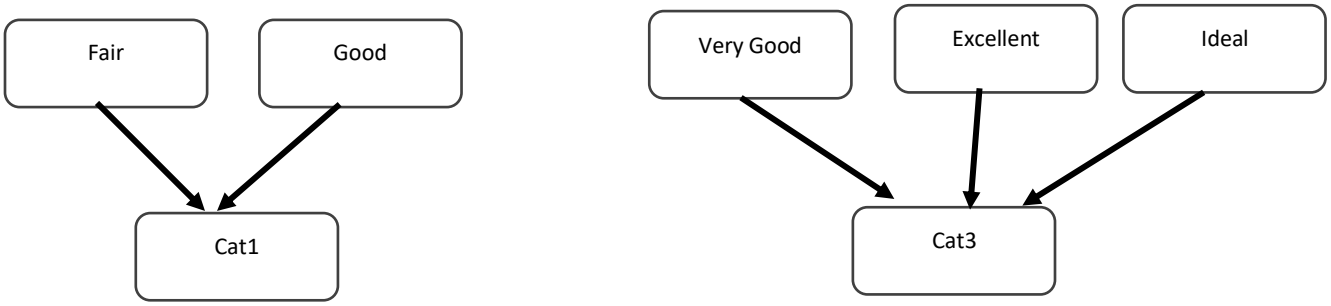
a. Dependent Variable: Price

b. Predictors: (Constant), CutN=Ideal, CutN=Very Good, CutN=good, CutN=Excellent

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2559.839	45.807		55.883	.000
	CutN=good	190.925	74.527	.182	2.562	.011
	CutN=Very Good	442.679	80.314	.382	5.512	.000
	CutN=Excellent	179.674	60.040	.230	2.993	.003
	CutN=Ideal	331.694	68.626	.353	4.833	.000

From the above result, cut “Good” are highly insignificant in determining price of diamond. As a further analysis cut is grouped into two categories as below.



Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.265[a]	.070	.066	$355.01480	1.024

a. Predictors: (Constant), CutN_2cat=2.0

b. Dependent Variable: Price

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2261113.690	1	2261113.690	17.940	.000[b]
	Residual	29996450.16	238	126035.505		
	Total	32257563.85	239			

a. Dependent Variable: Price

b. Predictors: (Constant), CutN_2cat=2.0

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2832.460	28.987		97.715	.000
	CutN_2cat=2.0	-200.493	47.335	-.265	-4.236	.000

a. Dependent Variable: Price

The above result shows model for cut with two categories (cat1.0 and cat2.0). Now, Cut has significant effect on price of diamond.

Polish vs Price

Here Polish is independent variable and price is dependent variable. By applying these variables in the regression model we got the results as shown in below snap.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.386[a]	.149	.134	$341.83556	1.421

a. Predictors: (Constant), PolishN=Ideal, PolishN=Excellent, PolishN=Very Good, PolishN=good

b. Dependent Variable: Price

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4797449.547	4	1199362.387	10.264	.000[b]
	Residual	27460114.30	235	116851.550		
	Total	32257563.85	239			

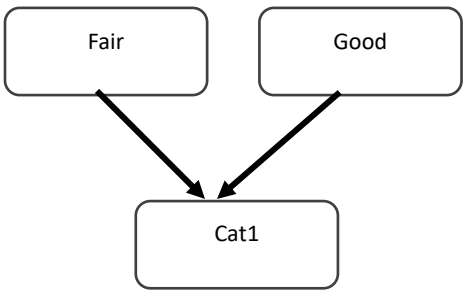
a. Dependent Variable: Price

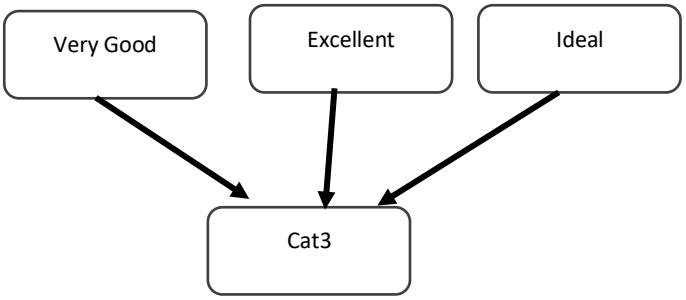
b. Predictors: (Constant), PolishN=Ideal, PolishN=Excellent, PolishN=Very Good, PolishN=good

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2318.600	152.874		15.167	.000
	PolishN=good	325.195	156.249	.443	2.081	.038
	PolishN=Very Good	524.318	156.764	.702	3.345	.001
	PolishN=Excellent	683.686	170.102	.527	4.019	.000
	PolishN=Ideal	728.800	216.196	.284	3.371	.001

From the above result, polish “Good” are highly insignificant in determining price of diamond. As a further analysis polish is grouped into two categories as below.





Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.339[a]	.115	.111	$346.37077	1.293

a. Predictors: (Constant), PolishN_2cat=2.0

b. Dependent Variable: Price

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3704058.788	1	3704058.788	30.874	.000[b]
	Residual	28553505.06	238	119972.710		
	Total	32257563.85	239			

a. Dependent Variable: Price

b. Predictors: (Constant), PolishN_2cat=2.0

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	2878.439	31.231		92.165	.000
	PolishN_2cat=2.0	-248.542	44.730	-.339	-5.556	.000

a. Dependent Variable: Price

The above result shows model for polish with two categories (cat1.0 and cat2.0). Now, polish has significant effect on price of diamond.

Certification vs Price

Here certification is independent variable and price is dependent variable. By applying these variables in the regression model we got the results as shown in below snap.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.225[a]	.051	.047	$356.46798	.957

a. Predictors: (Constant), CertificationN_2cat=2.0

b. Dependent Variable: Price

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1612429.234	1	1612429.234	12.689	.000[b]
	Residual	30115452.23	237	127069.419		
	Total	31727881.46	238			

a. Dependent Variable: Price

b. Predictors: (Constant), CertificationN_2cat=2.0

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	2842.108	32.541		87.340	.000
	CertificationN_2cat=2.0	-164.276	46.116	-.225	-3.562	.000

a. Dependent Variable: Price

From the above result, Certification has significant effect on price of diamond.

Symmetry vs Price

Here symmetry is independent variable and price is dependent variable. By applying these variables in the regression model we got the results as shown in below snap.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.375[a]	.141	.126	$343.39332	1.526

a. Predictors: (Constant), SymmetryN=Ideal, SymmetryN=Excellent, SymmetryN=Very Good, SymmetryN=good

b. Dependent Variable: Price

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4546605.274	4	1136651.318	9.639	.000[b]
	Residual	27710958.58	235	117918.973		
	Total	32257563.85	239			

a. Dependent Variable: Price

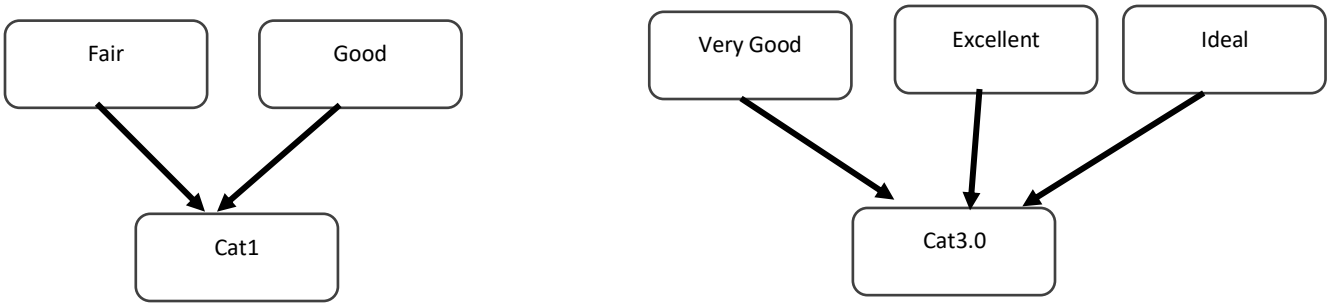
b. Predictors: (Constant), SymmetryN=Ideal, SymmetryN=Excellent, SymmetryN=Very Good, SymmetryN=good

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	2432.286	74.935		32.459	.000
	SymmetryN=good	260.791	82.152	.352	3.174	.002
	SymmetryN=Very Good	413.393	83.779	.538	4.934	.000
	SymmetryN=Excellent	502.868	100.750	.426	4.991	.000
	SymmetryN=Ideal	615.114	170.877	.240	3.600	.000

a. Dependent Variable: Price

From the above result, symmetry “Good” are highly insignificant in determining price of diamond. As a further analysis symmetry is grouped into two categories as below.



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.307[a]	.094	.091	$350.35493

a. Predictors: (Constant), SymmetryN_2cat=3.0

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3043402.466	1	3043402.466	24.794	.000[b]
	Residual	29214161.38	238	122748.577		
	Total	32257563.85	239			

a. Dependent Variable: Price

b. Predictors: (Constant), SymmetryN_2cat=3.0

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2649.264	31.337		84.542	.000
	SymmetryN_2cat=3.0	225.414	45.270	.307	4.979	.000

a. Dependent Variable: Price

The above result shows model for symmetry with two categories (cat1.0 and cat3.0). Now, symmetry has significant effect on price of diamond.

Multicollinearity

Below snap shows the correlation between the independent variables. Clarity and Carat are negatively correlated and all other variables are corelated in some may but, they are not highly correlated.

Correlations

		Carat	ColourN	ClarityN	CutN	PolishN	SymmetryN	CertificationN
Carat	Pearson Correlation	1	-.339**	-.632**	-.129*	-.392**	-.376**	-.215**
	Sig. (2-tailed)		.000	.000	.047	.000	.000	.001
	N	240	240	239	240	240	240	239
ColourN	Pearson Correlation	-.339**	1	-.163*	.031	.058	.095	.114
	Sig. (2-tailed)	.000		.012	.632	.368	.143	.079
	N	240	240	239	240	240	240	239
ClarityN	Pearson Correlation	-.632**	-.163*	1	.139*	.425**	.417**	.043
	Sig. (2-tailed)	.000	.012		.032	.000	.000	.507
	N	239	239	239	239	239	239	238
CutN	Pearson Correlation	-.129*	.031	.139*	1	.421**	.521**	.147*
	Sig. (2-tailed)	.047	.632	.032		.000	.000	.023
	N	240	240	239	240	240	240	239
PolishN	Pearson Correlation	-.392**	.058	.425**	.421**	1	.685**	.208**
	Sig. (2-tailed)	.000	.368	.000	.000		.000	.001
	N	240	240	239	240	240	240	239
SymmetryN	Pearson Correlation	-.376**	.095	.417**	.521**	.685**	1	.040
	Sig. (2-tailed)	.000	.143	.000	.000	.000		.542
	N	240	240	239	240	240	240	239
CertificationN	Pearson Correlation	-.215**	.114	.043	.147*	.208**	.040	1
	Sig. (2-tailed)	.001	.079	.507	.023	.001	.542	
	N	239	239	238	239	239	239	239

**. Correlation is significant at the 0.01 level (2-tailed).

Final Model with All above variables

From the above analysis, our model will consist of below independent variables.

- ❑ **Colour2cat** - Colourless vs Coloured
- ❑ **Clarity2cat** - Visible, 30x-10x
- ❑ **Cut2catN** -Fair-Good, Very Good- Excellent-Ideal
- ❑ **PolishN_2cat** - Fair-Good, Very Good-Excellent-Ideal
- ❑ **Symmetry_3cat** – Fair-good, Very Good-Excellent-Ideal

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.652[a]	.425	.408	$281.00578	1.994

a. Predictors: (Constant), SymmetryN_2cat=3.0, Colour_2Cat=2.0, CertificationN_2cat=3.0, Clarity2cat=2.0, CutN_2cat=3.0, PolishN_2cat=3.0, Carat

b. Dependent Variable: Price

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13487139.82	7	1926734.260	24.400	.000[b]
	Residual	18240741.65	231	78964.250		
	Total	31727881.46	238			

a. Dependent Variable: Price

b. Predictors: (Constant), SymmetryN_2cat=3.0, Colour_2Cat=2.0, CertificationN_2cat=3.0, Clarity2cat=2.0, CutN_2cat=3.0, PolishN_2cat=3.0, Carat

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	1431.173	260.445		5.495	.000
	Carat	711.320	215.761	.238	3.297	.001
	Colour_2Cat=2.0	259.783	47.097	.335	5.516	.000
	CertificationN_2cat=3.0	101.418	39.659	.139	2.557	.011
	PolishN_2cat=3.0	111.500	43.876	.153	2.541	.012
	Clarity2cat=2.0	465.459	49.762	.635	9.354	.000
	CutN_2cat=3.0	90.803	41.727	.120	2.176	.031
	SymmetryN_2cat=3.0	37.132	45.051	.051	.824	.411

a. Dependent Variable: Price

In the above Multi linear regression model Symmetry is highly insignificant. So, We reran the regression including all variables except Symmetry.

Final Model

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.651[a]	.423	.408	$280.81153	1.980

a. Predictors: (Constant), CutN_2cat=3.0, Colour_2Cat=2.0, Clarity2cat=2.0, CertificationN_2cat=3.0, PolishN_2cat=3.0, Carat

b. Dependent Variable: Price

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1453.443	258.861		5.615	.000
	Carat	692.103	214.349	.232	3.229	.001
	Colour_2Cat=2.0	261.537	47.016	.338	5.563	.000
	CertificationN_2cat=3.0	98.790	39.503	.136	2.501	.013
	PolishN_2cat=3.0	126.375	39.965	.173	3.162	.002
	Clarity2cat=2.0	468.827	49.560	.639	9.460	.000
	CutN_2cat=3.0	99.527	40.334	.132	2.468	.014

a. Dependent Variable: Price

Our final model contains following independent variables.

- **Colour2cat** - Colourless vs Coloured
- **Clarity2cat** - Visible, 30x-10x
- **Cut2catN** -Fair-Good, Very Good- Excellent-Ideal
- **PolishN_2cat** - Fair-Good, Very Good-Excellent-Ideal

Our model explains 42.3% of variation.

Now substituting the values that professor looking for in our model will give the value of the diamond.

- Price = 1453.443+ 261.537*colour_2 + 468.827*clarity_2 + 99.527*Cut_2 + 692.103*carat +126.375*Polish_2 + 98.790*certification_2
- Price = 1453.443 + 261.537*0 + 468.827*1 + 99.527*1 + 692.103*0.9 + 126.375*0 + 98.790*1

Finally, Fair Price of the diamond Comes out to be $2643.

Summary of our model

As per our model the diamond price that professor should be paying is $2643. Our model does not include the price of the ring. If final price of diamond and ring is around $3100 then, professor can go ahead in buying diamond ring.

Conclusion

- Color, Clarity, Cut and Polish determines the price of the diamond.
- Our model determines the Price of diamond, not the price of ring.
- Our model is valid for the given data not for other variables.
- Since professor is buying diamond ring once and assuming price of ring would be around $300-$400. Total price is 3043, Professor can go ahead and buy the ring.