The Professor Proposes

MCDA 5520 Team Project

Group Members: Caner Irfanoglu
Diven Sambhwani
Madeleine Leong
Sreeraj Punnoli
Tom Tong

December 4th, 2018

CONTENTS

01

Instruction &
Problem Statement

02

Descriptive
Analysis

03

Pre-work

04

Feature
Engineering

05

Model
Selection

06

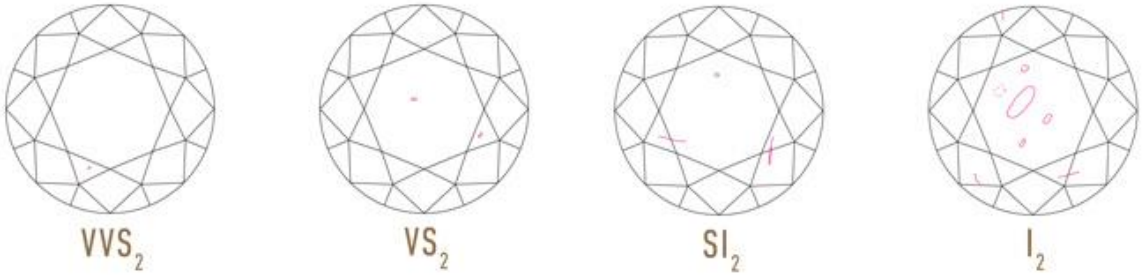
Conclusion

01

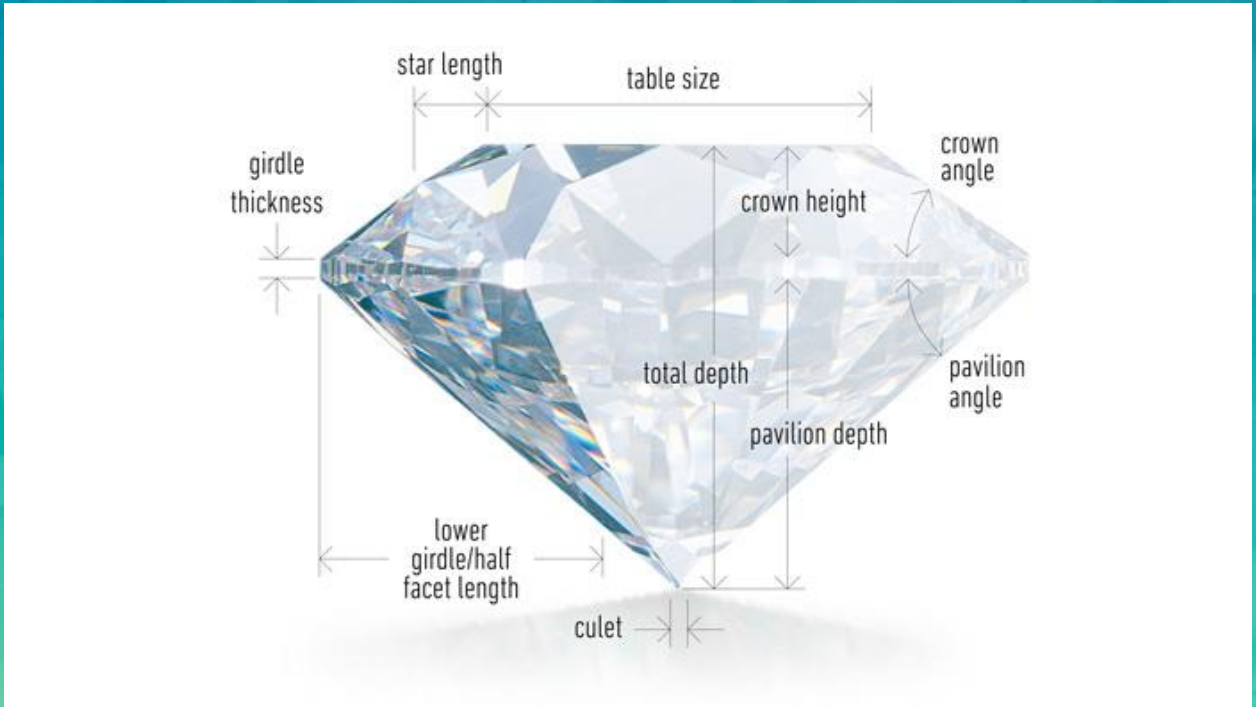
Instruction & Problem Statement

Color

Clarity



Cut


star length
table size
girdle
thickness
crown
angle
crown height
total depth
pavilion
angle
pavilion depth
lower
girdle/half
facet length
culet

Carat

Problem Statement

GIA®

FACSIMILE

This is a digital representation of the original GIA Report. This representation might not be accepted in lieu of the original GIA Report in certain circumstances. The original GIA Report includes certain security features which are not reproducible on this facsimile.

GIA REPORT
2287173890

Verify this report at GIA.edu

GIA NATURAL DIAMOND DOSSIER®

February 17, 2018

GIA Report Number	2287173890
Shape and Cutting Style	Round Brilliant
Measurements	6.06 - 6.11 x 3.89 mm

GRADING RESULTS

Carat Weight	0.90 carat
Color Grade	J
Clarity Grade	SI2
Cut Grade	Very Good

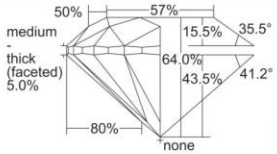
ADDITIONAL GRADING INFORMATION

Polish	Very Good
Symmetry	Very Good
Fluorescence	None
Clarity Characteristics	Feather

Inscription(s): GIA 2287173890

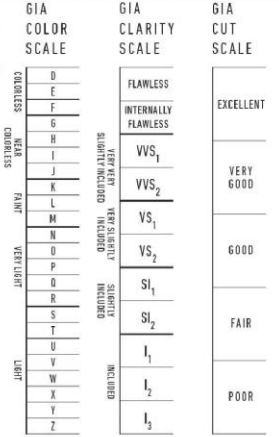
GIA.edu

PROPORTIONS



Profile to actual proportions

GRADING SCALES



The results documented in this report refer only to the diamond described, and were obtained using the techniques and equipment available to GIA at the time of examination. This report is not a guarantee or valuation. For additional information and important limitations and disclaimers, please see GIA.edu/terms or call +1 800 421 7250 or +1 760 603 4500. ©2017 Gemological Institute of America, Inc.

THE SECURITY FEATURES IN THIS DOCUMENT, INCLUDING THE HOLOGRAM, SECURITY SCREEN AND MICROPRINT LINES, IN ADDITION TO THOSE NOT LISTED, EXCEED DOCUMENT SECURITY INDUSTRY GUIDELINES.

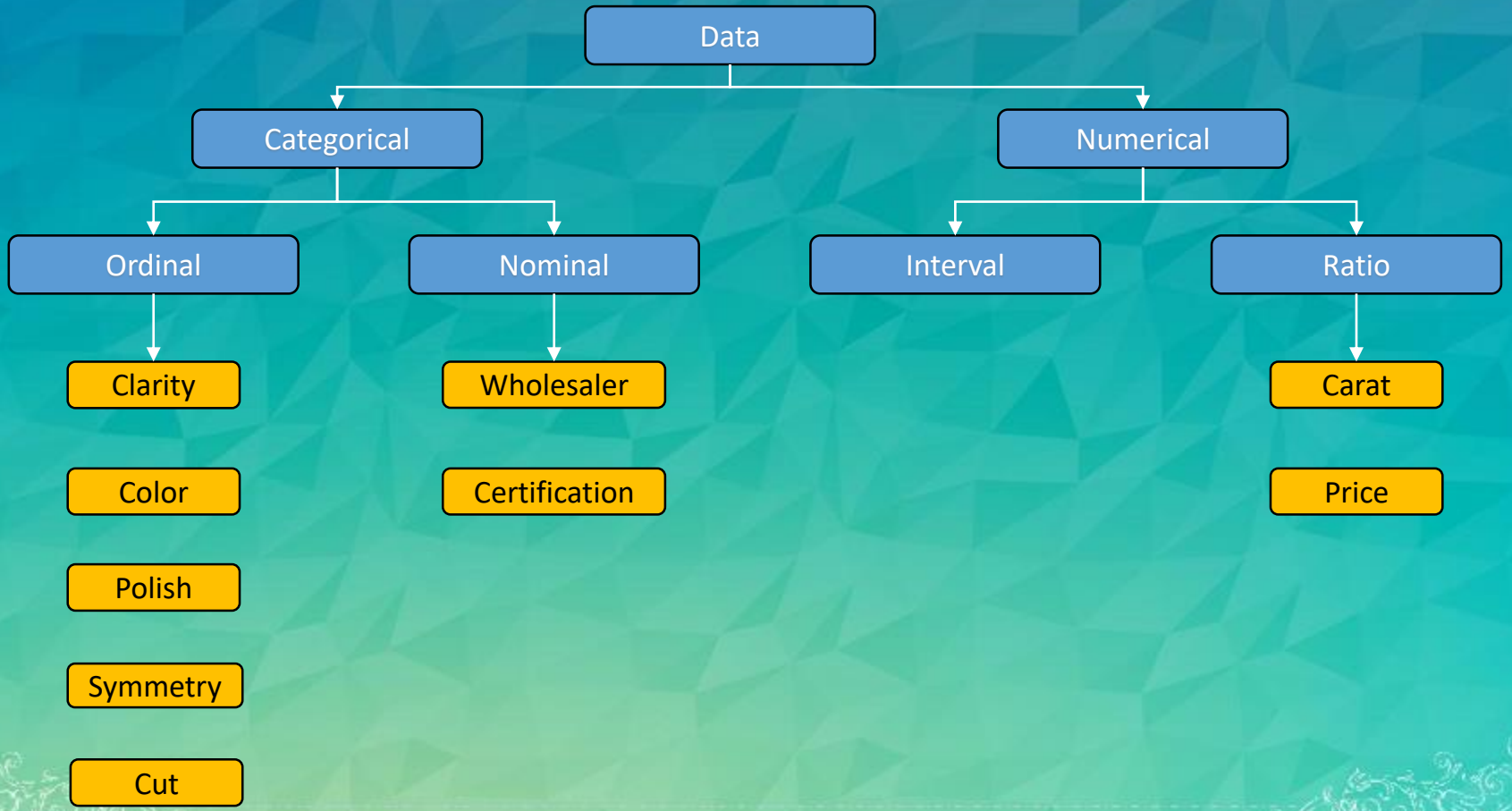
reportcheck.gia.edu

2

02

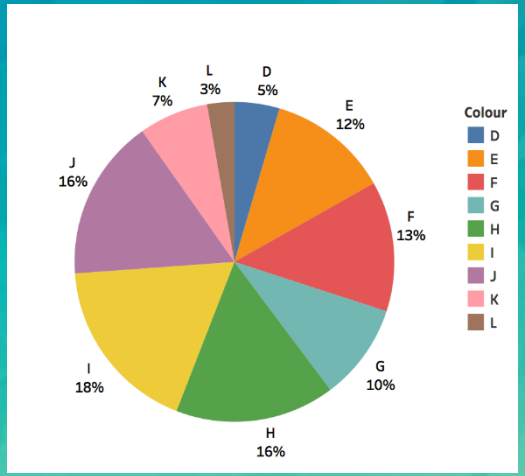
Descriptive Analysis

Data Types of the Variables

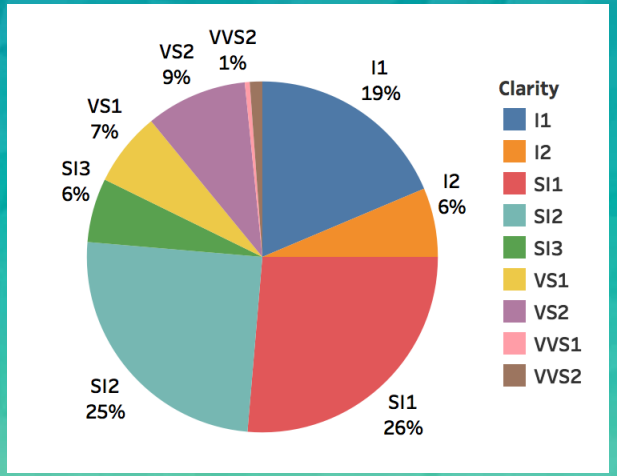


Descriptive Analysis

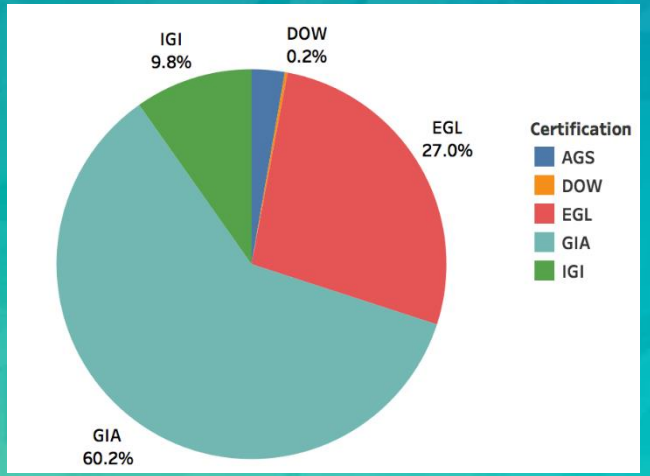
Distribution of Independent Variables – I



Color



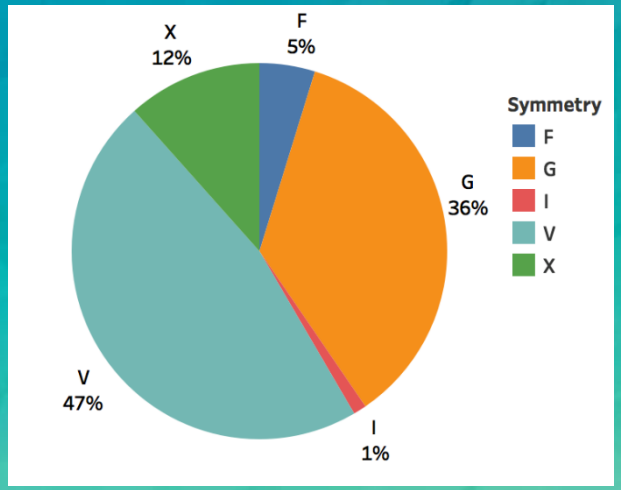
Clarity



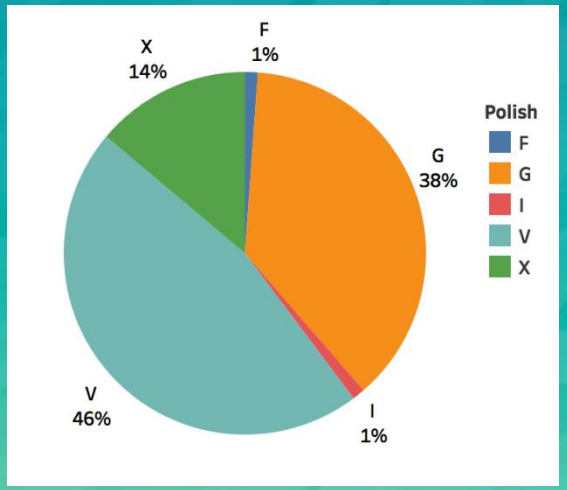
Certification

Descriptive Analysis

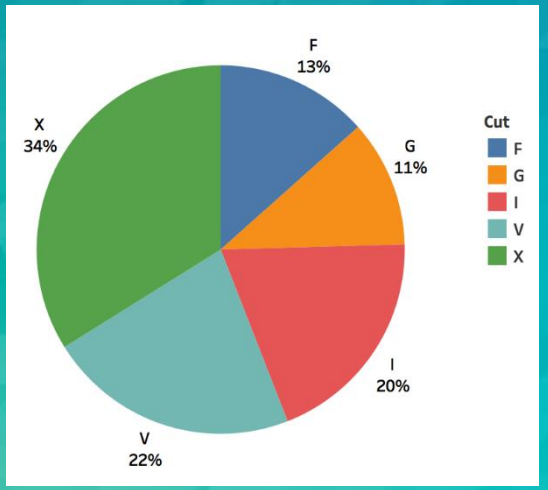
Distribution of Independent Variables – II



Symmetry



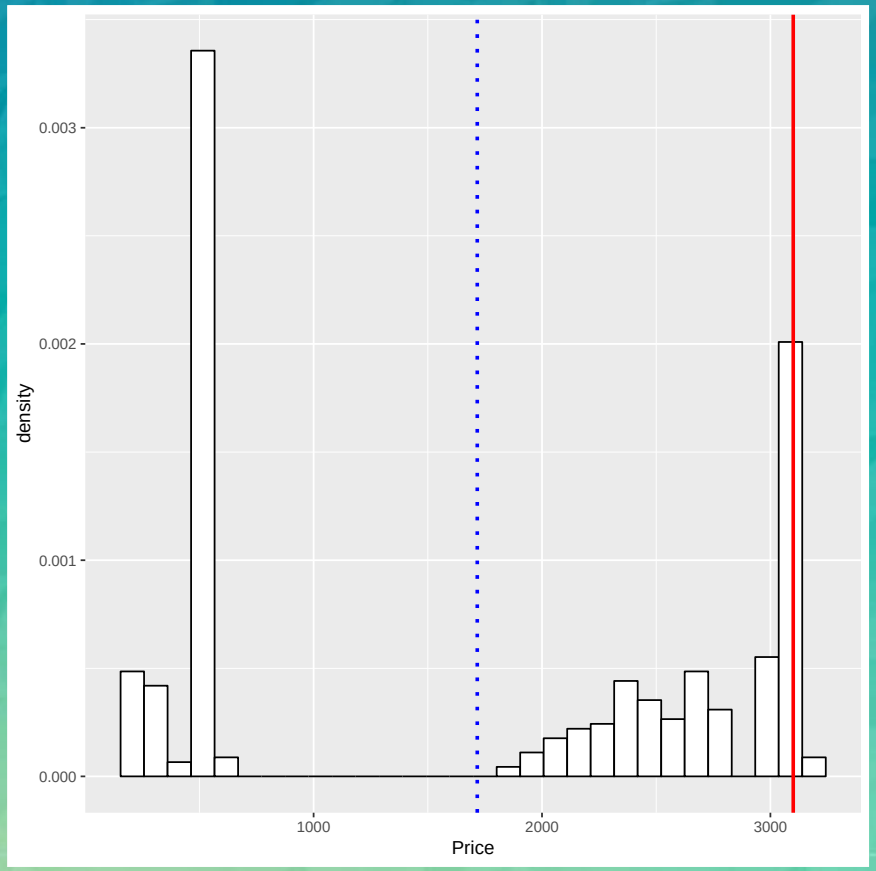
Polish



Cut

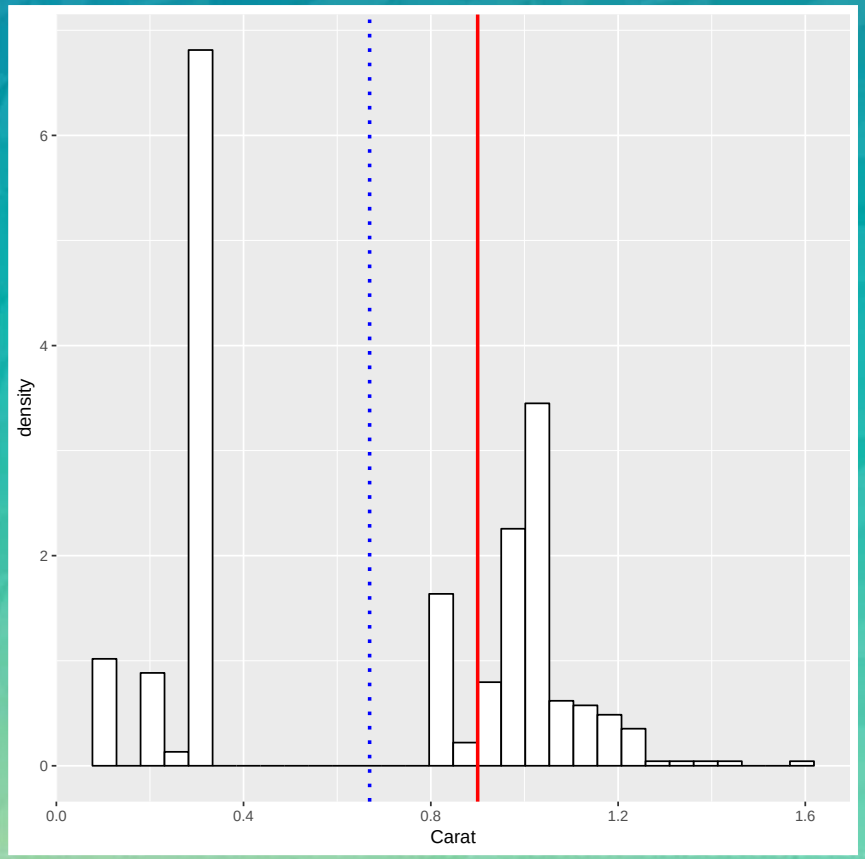
Descriptive Analysis

Density Distribution for Price



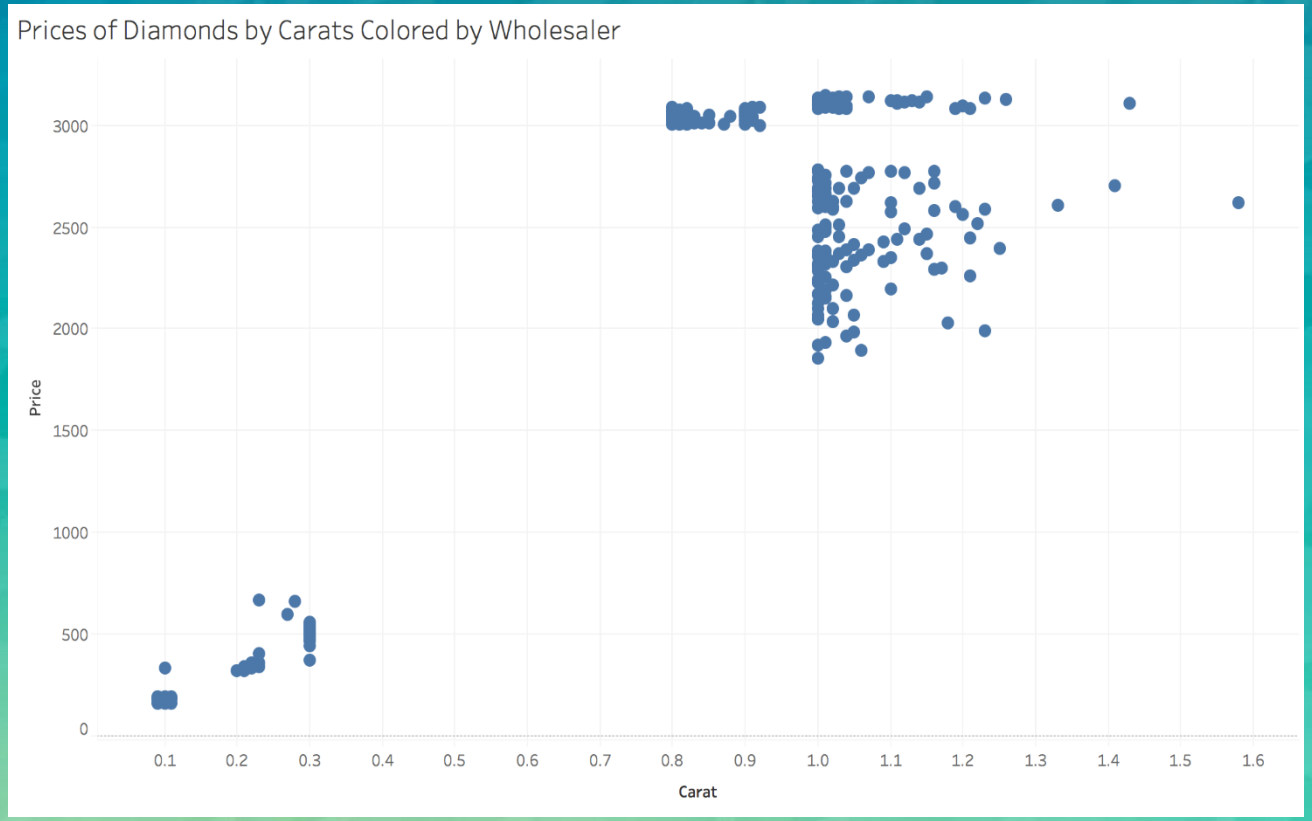
Descriptive Analysis

Density Distribution for Price



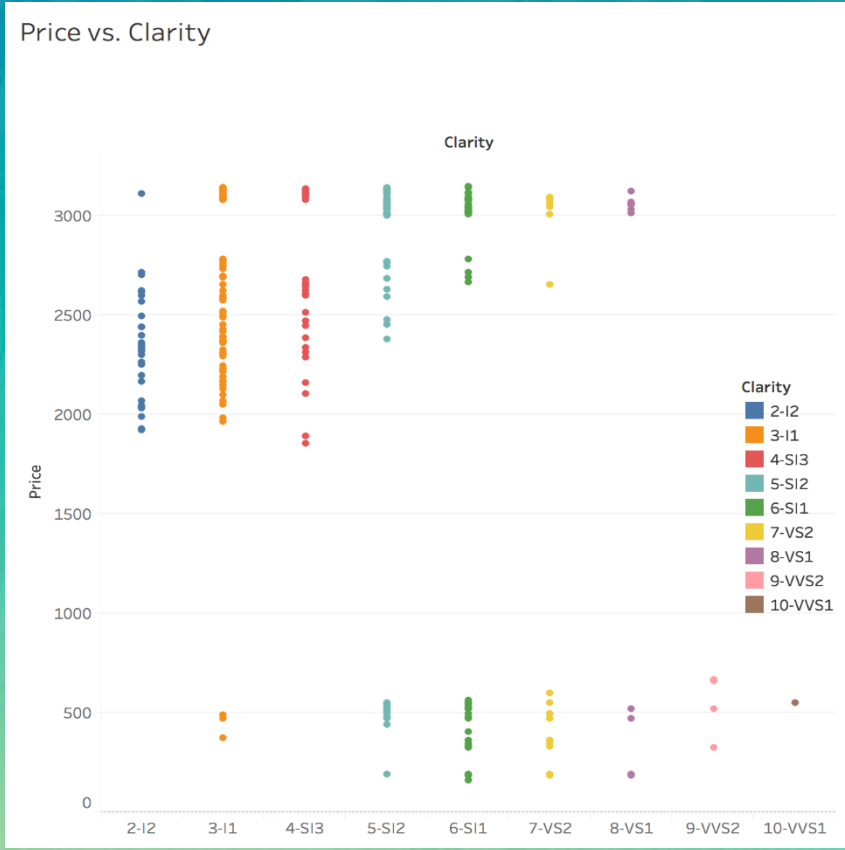
Descriptive Analysis

Bivariates: Price vs Carat



Descriptive Analysis

Bivariates: Price vs. Clarity



Descriptive Analysis

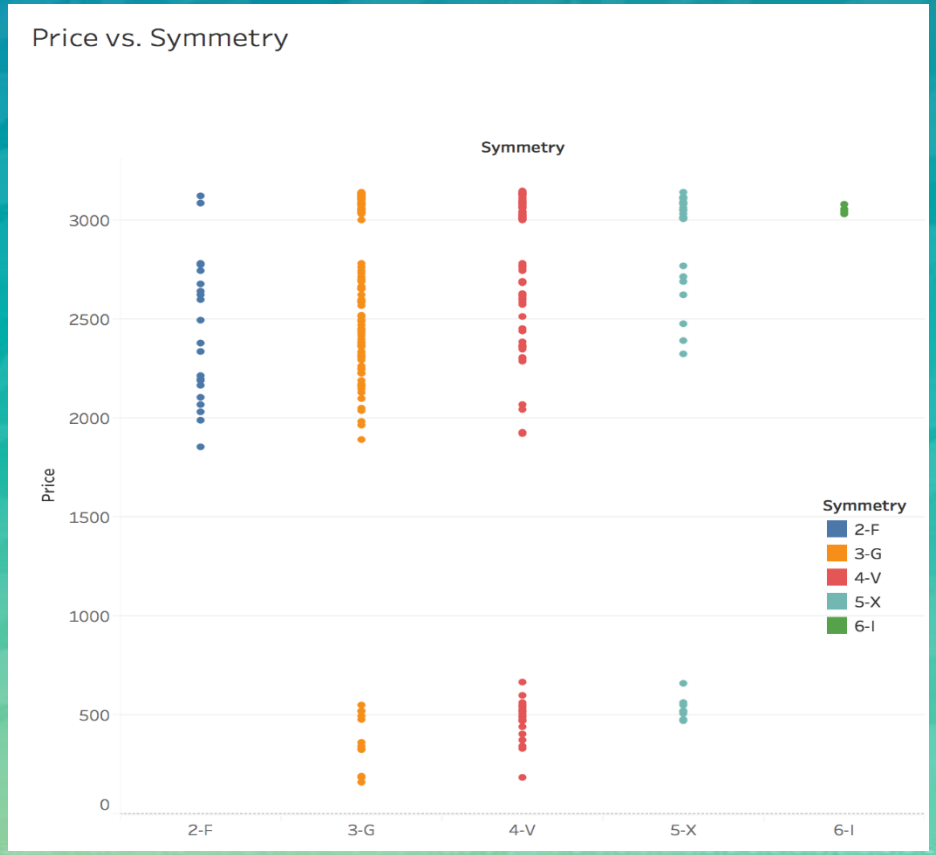
Bivariates: Price vs Cut

Descriptive Analysis

Bivariates: Price vs Polish

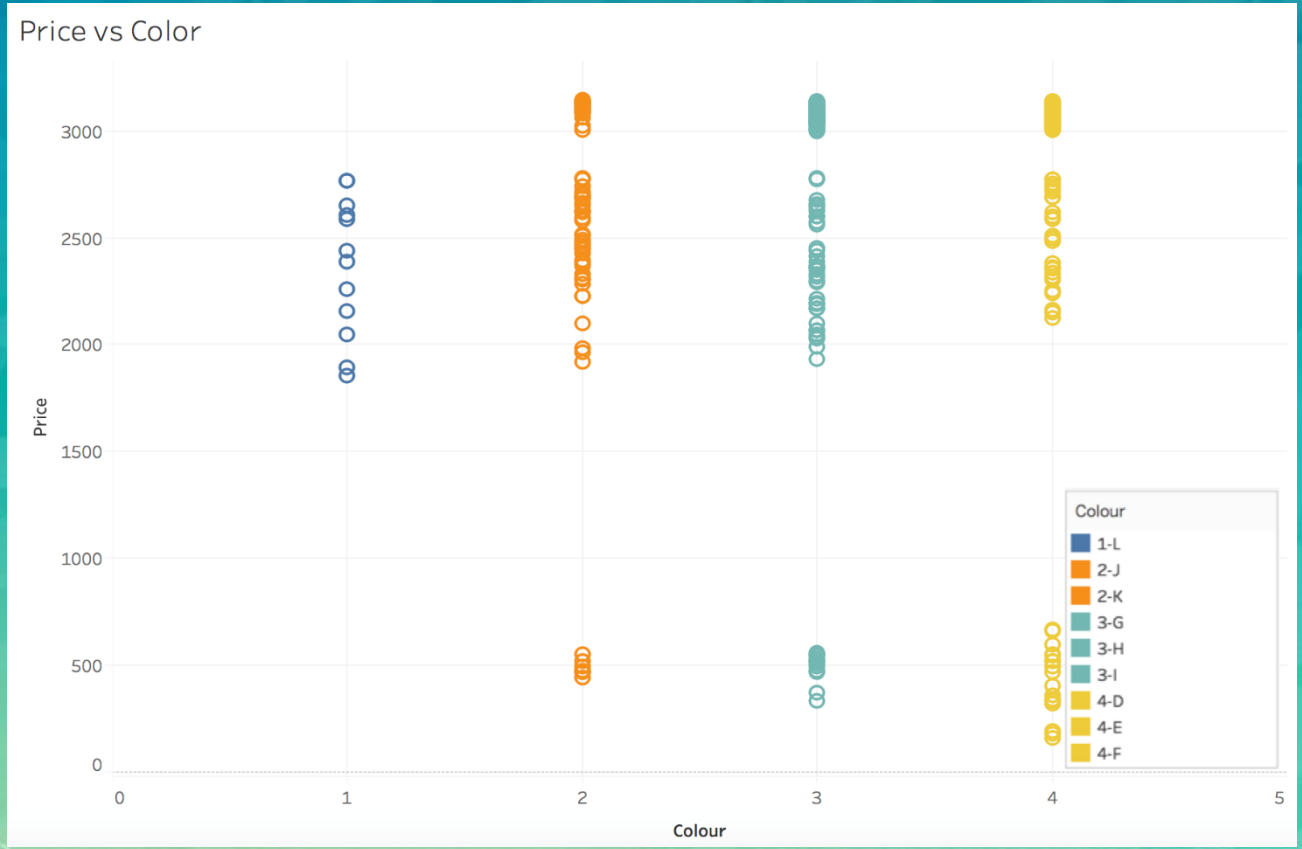
Descriptive Analysis

Bivariates: Price vs Symmetry



Descriptive Analysis

Bivariates: Price vs Color



Pre-work

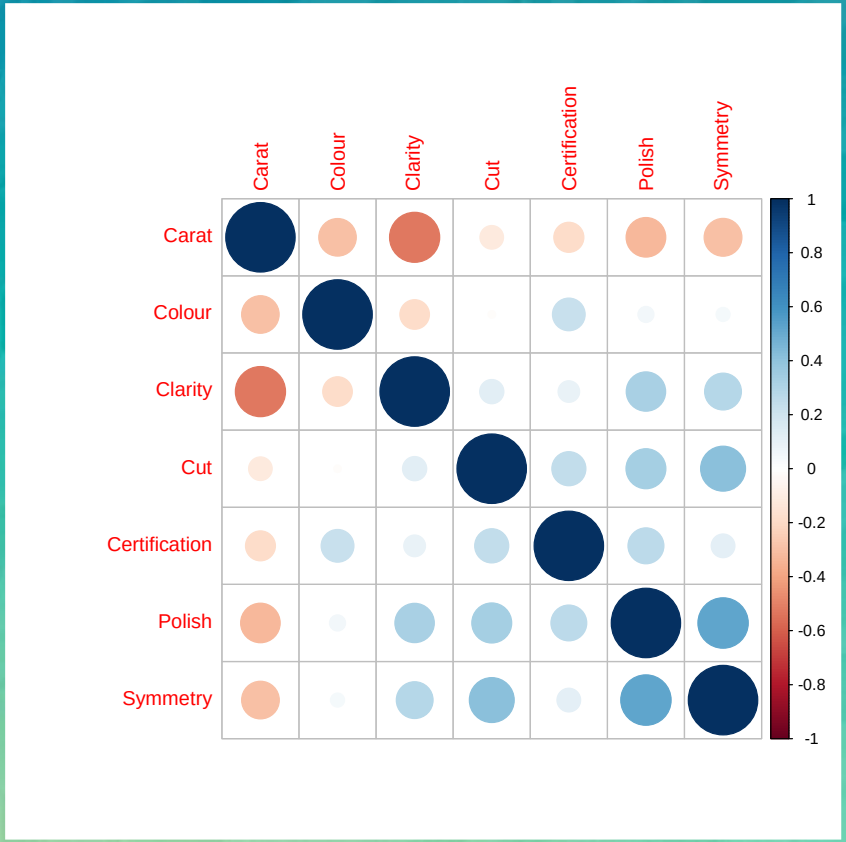
Pre-work

ANOVA

Variable v/s Price	F-Value	Critical Value	P-Value
Cut	17.94	2.41	0.000
Color	5.32	3.88	0.021
Symmetry	17.89	3.03	0.000
Polish	18.14	3.03	0.000
Clarity	42.64	3.03	0.000
Certification	4.08	3.88	0.044

Pre-work

MULTICOLLINEARITY



04

Feature Engineering

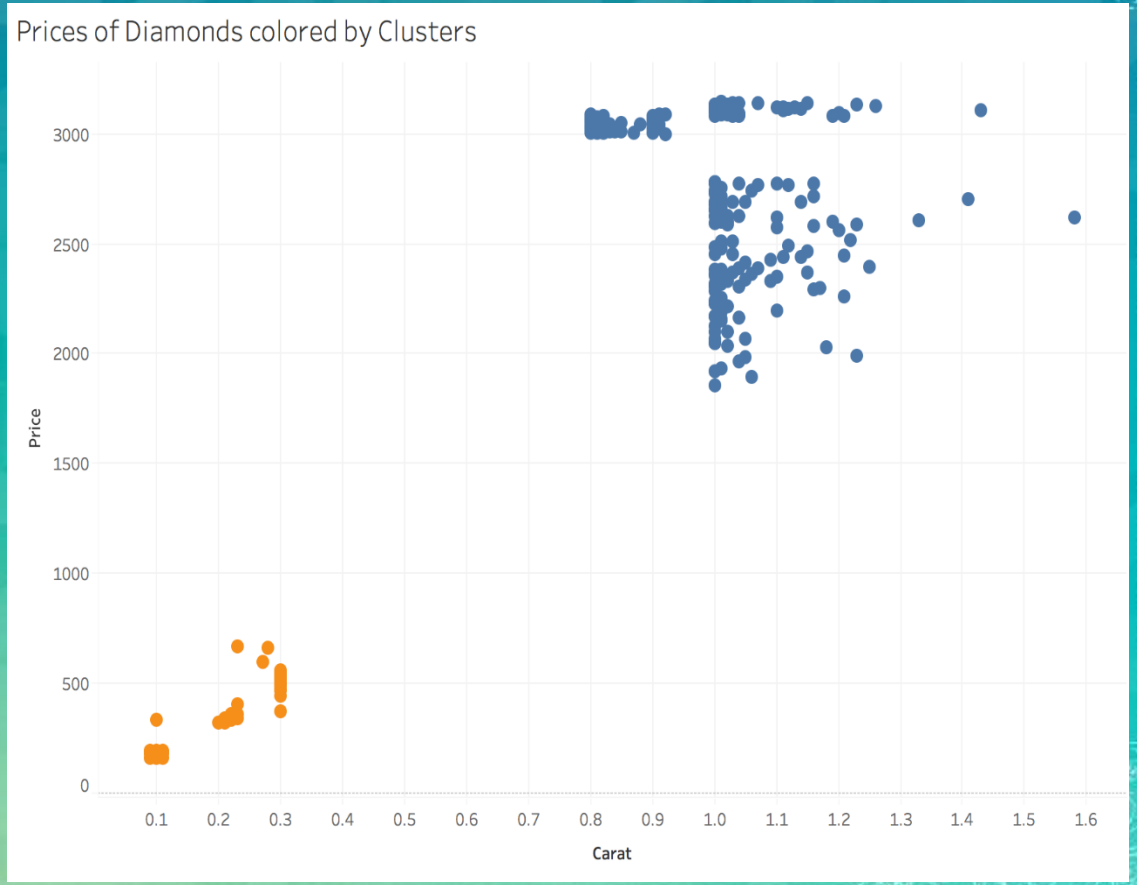
Why Performing Analysis on Blue Cluster

Carat is a significant predictor

Professor's Diamond lies in the blue cluster

Using all available data may lead bias, since Lower carat and Higher carat traits may differ

Most of the data for blue cluster lies between 0.8 - 1.3 carats, which is a relatively small interval Blue cluster is more favorable under linearity assumption (Carat vs. Price not linear for the whole carat range - as given in the pdf)



Feature Engineering Part I

Getting to know variables

Price is the prediction variable (Dependent)
Carat, Clarity, Color, Polish, Symmetry, Cut, Certification, Wholesaler are predictors (Independent variables)
Categorical Variable levels should be distinct based on their price prediction ability
Wholesaler needs to be excluded from prediction, since it is not a diamond characteristic (May introduce bias)

Feature Engineering Part II

Grouping Ordinal Variables by Predictive Ability

Clarity

Originally had 7 levels.
Grouped to 3 levels

Variable Name	Original Levels	Final Levels	Levelling Criteria(s)	Original R-squared	After ReGrouping R-squared
Clarity	I2 I1 SI3 SI2 SI1 VS2 VS1	Flawed Naked Eye 10x Zoom Flaws 30x Zoom Flaws	Some Levels Insignificant for price	0.403	0.265

Feature Engineering Part II

Grouping Ordinal Variables by Predictive Ability

Color

Variable Name	Original Levels	Final Levels	Levelling Criteria(s)	Original R-squared	After ReGrouping R-squared
Color	L J,K G,H,I F,D,E	Near Colorless Lightly Yellow	Some Levels Insignificant for price	0.065	0.021

Feature Engineering Part II

Grouping Ordinal Variables by Predictive Ability

Polish

Variable Name	Original Levels	Final Levels	Levelling Criteria(s)	Original R-squared	After ReGrouping R-squared
Polish	F G V X I	F + G V X + I	Small sample size for F and I	0.149	0.133

Feature Engineering Part II

Grouping Ordinal Variables by Predictive Ability

Symmetry

Variable Name	Original Levels	Final Levels	Levelling Criteria(s)	Original R-squared	After ReGrouping R-squared
Symmetry	F G V X I	F G V + X + I	• Small sample size for I • Low predictive ability difference between V-X	0.141	0.133

Feature Engineering Part II

Grouping Ordinal Variables by Predictive Ability

Cut

Variable Name	Original Levels	Final Levels	Levelling Criteria(s)	Original R-squared	After ReGrouping R-squared
Cut	F G V X I	F G V X I	• All levels distinct • Bin Sizes are large enough	0.144	0.133

Feature Engineering Part II

Grouping Ordinal Variables by Predictive Ability

Certification

Variable Name	Original Levels	Final Levels	Levelling Criteria(s)	Original R-squared	After ReGrouping R-squared
Certification	AGS GIA EGL DOW IGI	AGS + GIA EGL + DOW + IGI	Two most respected labs vs. others	0.082	0.054

05

Model Selection

Model Selection

Variable Selection

- Step 1: Construct a model with all independent variables
  - Cut variable is partially significant
- Regroup Cut into 2 groups
  - Fair, Good = 1
  - Very Good, Exceptional, Ideal = 2

```
Call:
lm(formula = Price ~ ., data = professor_cluster)

Residuals:
    Min      1Q  Median      3Q     Max
-777.74 -152.99   -3.54  149.63  683.01

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     1189.05     281.80   4.219 3.55e-05 ***
Carat            776.36     231.59   3.352 0.000939 ***
Colour2          265.62      47.20   5.628 5.37e-08 ***
Clarity2         467.67      50.25   9.307  < 2e-16 ***
Clarity3         557.34      94.33   5.908 1.26e-08 ***
Cut3              51.87      63.50   0.817 0.414862
Cut4             115.11      79.32   1.451 0.148084
Cut5             105.61      54.89   1.924 0.055595 .
Cut6              44.53      67.46   0.660 0.509880
Certification2    89.73      41.25   2.175 0.030652 *
Polish2          102.61      45.05   2.277 0.023695 *
Polish3          141.98      76.86   1.847 0.066007 .
Symmetry2        190.08      71.35   2.664 0.008273 **
Symmetry3        207.44      80.75   2.569 0.010845 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 278.2 on 226 degrees of freedom
Multiple R-squared:  0.4577,    Adjusted R-squared:  0.4265
```

Model Selection

Variable Selection

- Adjusted R^2 increased from 42.65% to 42.88%
- Take consideration for Cut and Polish variables because the p-value is approx. to 10%

```
Call:
lm(formula = Price ~ ., data = professor_cluster)

Residuals:
    Min      1Q  Median      3Q     Max
-802.25 -161.47   -8.02  160.39  683.03

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     1187.63     270.03   4.398 1.67e-05 ***
Carat            782.85     216.62   3.614 0.000371 ***
Colour2          262.90      46.63   5.639 5.02e-08 ***
Clarity2         467.22      49.34   9.469  < 2e-16 ***
Clarity3         553.02      93.35   5.924 1.14e-08 ***
Cut2              68.91      42.26   1.631 0.104323
Certification2   100.21      39.99   2.506 0.012908 *
Polish2           96.89      44.67   2.169 0.031109 *
Polish3          115.84      73.58   1.574 0.116797
Symmetry2        209.59      69.22   3.028 0.002745 **
Symmetry3        225.33      78.21   2.881 0.004338 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 277.7 on 229 degrees of freedom
Multiple R-squared:  0.4527,    Adjusted R-squared:  0.4288
F-statistic: 18.94 on 10 and 229 DF,  p-value: < 2.2e-16
```

Conclusion

Conclusion

Final Pricing Model for Diamond

- Price for professor's diamond ring = $2,753.87
- Difference from the quoted price = 346.14
- Consider the ring material

Variable	Coefficient	
Carat	782.85	
Colour2	262.90	D-I – 1 Else – 0
Clarity2	467.22	SI1, SI2,SI3 – 1 Else – 0
Clarity3	553.02	VS1, VS2, VVS1, VVS2 – 1 Else – 0
Cut2	68.91	Very Good, Exceptional, Ideal – 1 Else – 0
Certification2	100.21	AGS, GIA – 1 Else – 0
Polish2	96.89	Very Good – 1 Else – 0
Polish3	115.84	Exceptional, Ideal – 1 Else – 0
Symmetry2	209.59	Good – 1 Else – 0
Symmetry3	225.33	Very Good, Exceptional, Ideal – 1 Else – 0
Intercept	1187.63	

THANK YOU!