**Introduction & Problem statement –**

Keeping in mind that the diamond he buys for his girlfriend would suit her taste, as well as his budget of 2000$-4000$, the professor’s plans got collapsed, when he was confronted with an dizzying array of diamond characteristics, configuration and pricing.

With the information gained from the visit to mall, the professor wants to know whether a fair price was quoted for the diamond. So data is collected from three diamond wholesalers, he found on internet and he wants to combine them to derive the value for the diamond he found.

There are several characteristics that determine the cost of diamond:

* Carat
* Color
* Cut
* Clarity
* Polish
* Symmetry
* Certification

How are these Characteristics Classified:

* Carat – **0.1 – 1.58**
* Colour – **(Colourless) D-F | G-I | J-K | L-N | O-S | T-Z (Yellow)**
* Clarity – **(best)FL-IF | VVS1-VVS2 | VS1 –VS2 | SI1 – SI3 | I1 – I3 (least)**
* Cut
* Polish **Poor | Fair | Good | Very Good | Excellent | Ideal**
* Symmetry
* Certification – **(best)GIA | AGS | EGL | IGI**
* Wholesaler – **1 | 2 | 3**

Professor’s Diamond Characteristics:

* Carat weight : **0.9**
* Cut : **Very Good**
* Color : **J (Faint Yellow)**
* Clarity: **SI2(Few inclusions visible at 10x)**
* Polish - **Good**
* Symmetry - **Very Good**
* Certification: **GIA**

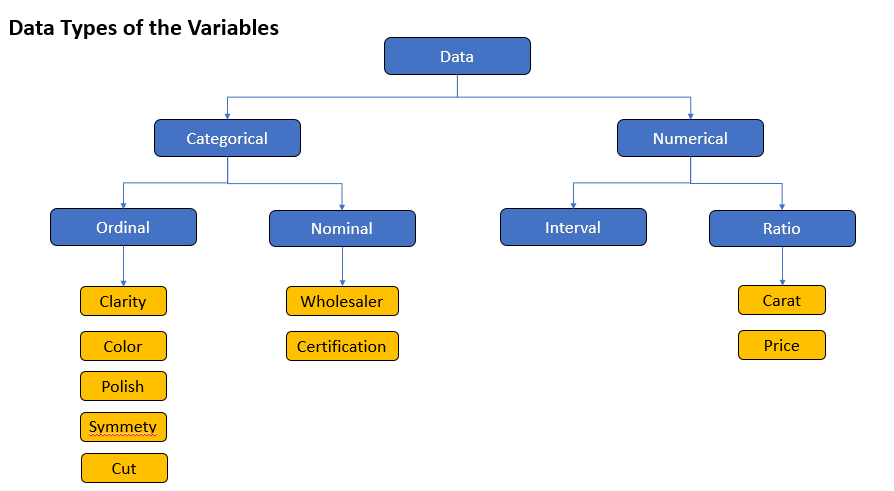
**QUOTED PRICE FOR THIS DIAMOND IS**

**$3100**

What should be the **Fair price** of the diamond?

**Cleaning the data**

Redraw this in some other way/different colours



PS : don’t copy this in report

Transforming the variables

clarity

i3 1

i2 2

i1 3

si3 4

si2 5

si1 6

vs2 7

vs18

vvs2 9

vvs1 10

Cut polish symmetry

P - 1

F - 2

G - 3

V - 4

X - 5

I - 6

Color

1- very light yellow

2-faint

3 – near colorless

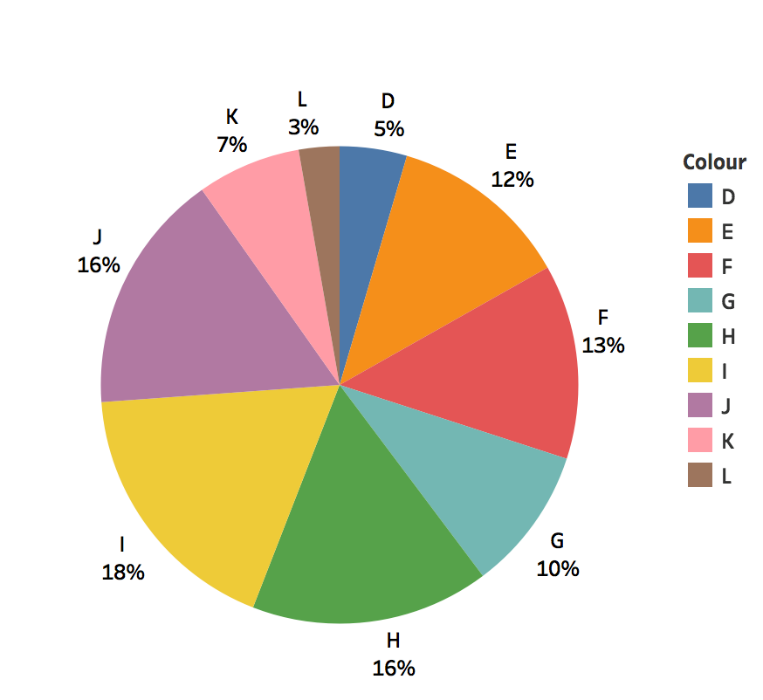
4- colorless

For Certification- Refer PPT

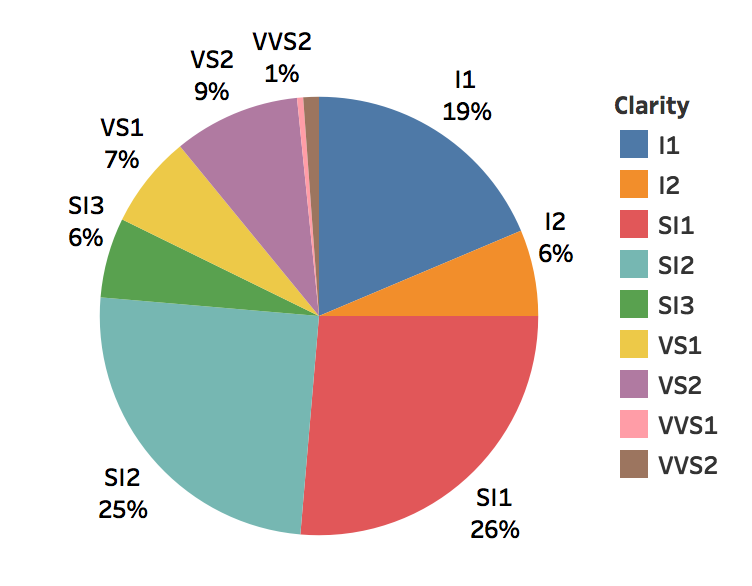
**Descriptive**

Take tabular screenshots from PPT and below

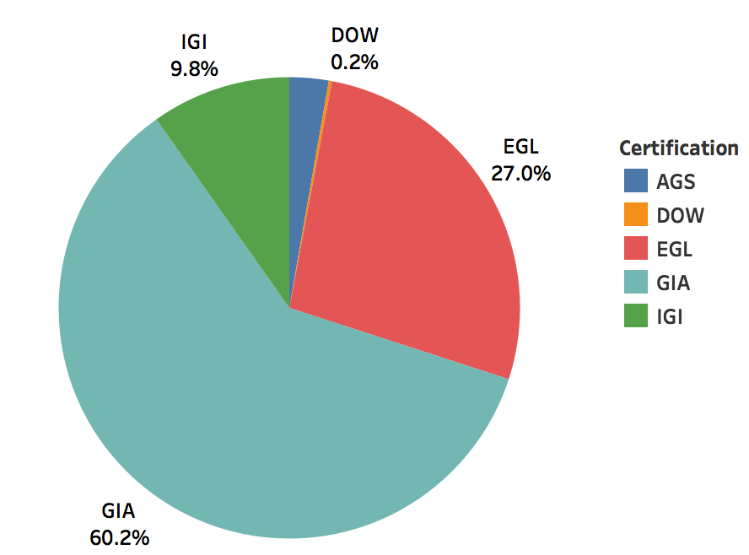
Colour



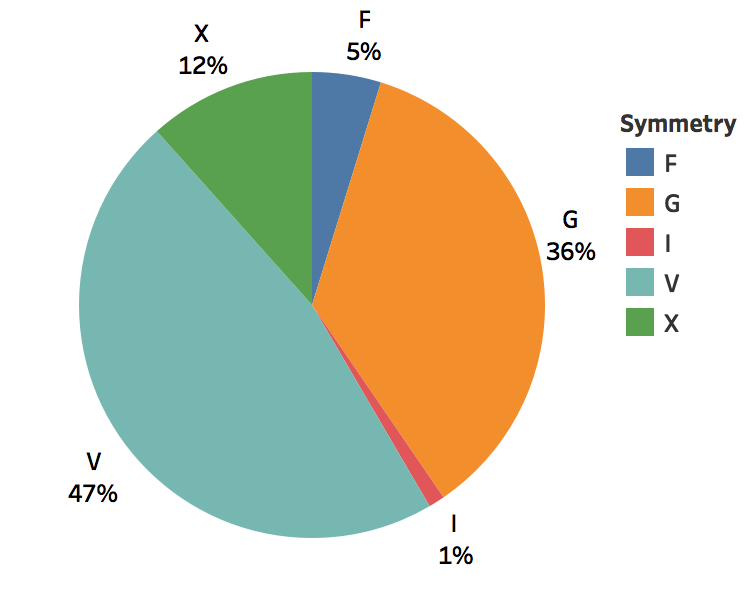
Clarity



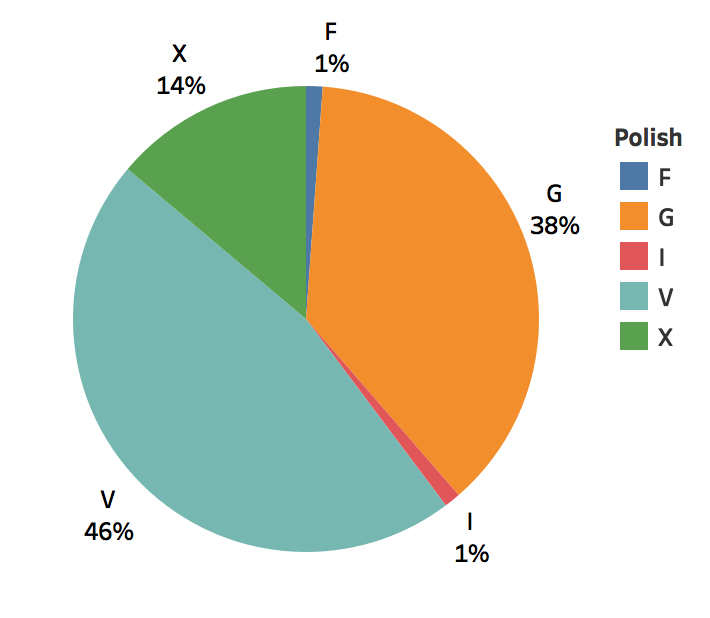
Certification



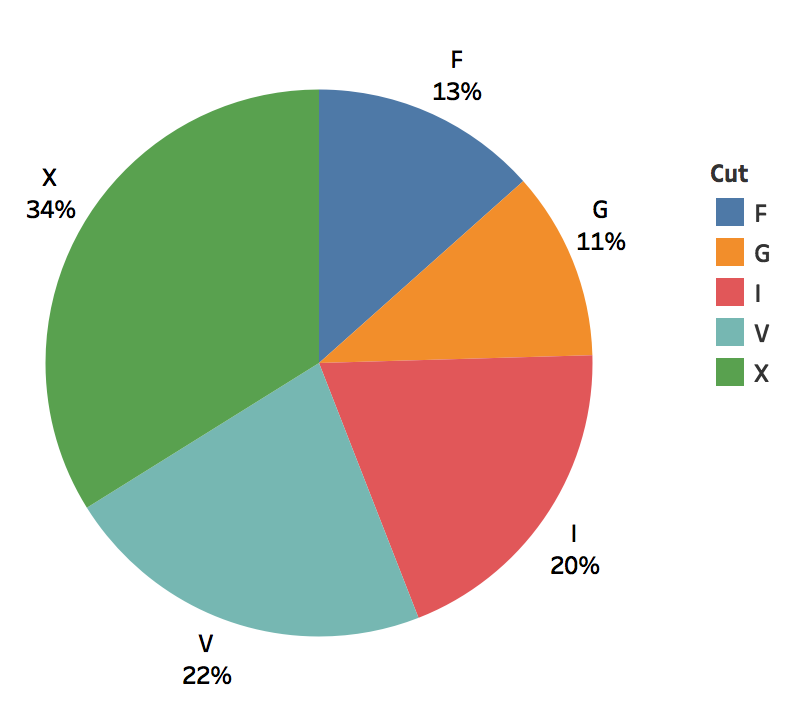
Symmetry



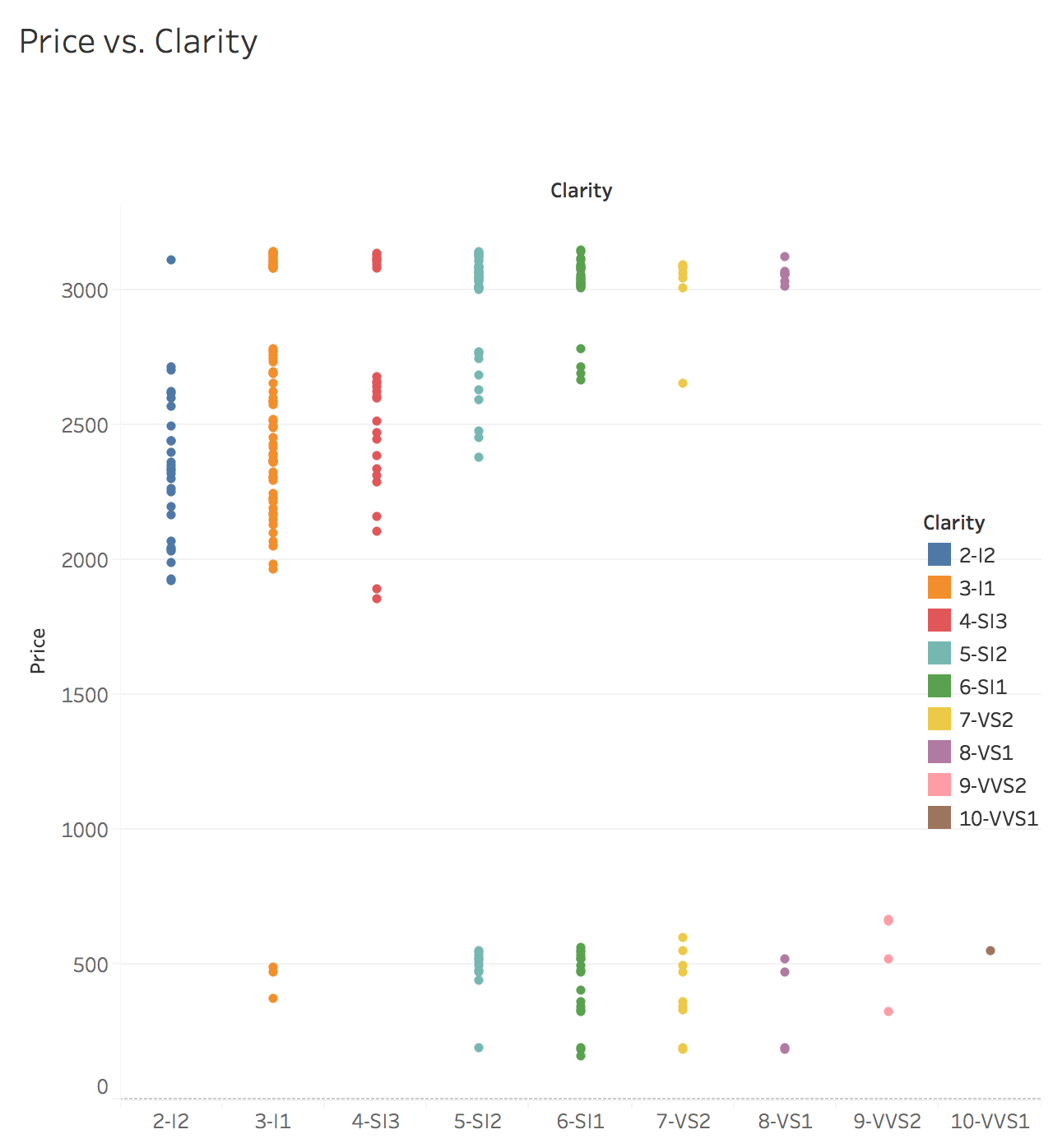
Polish

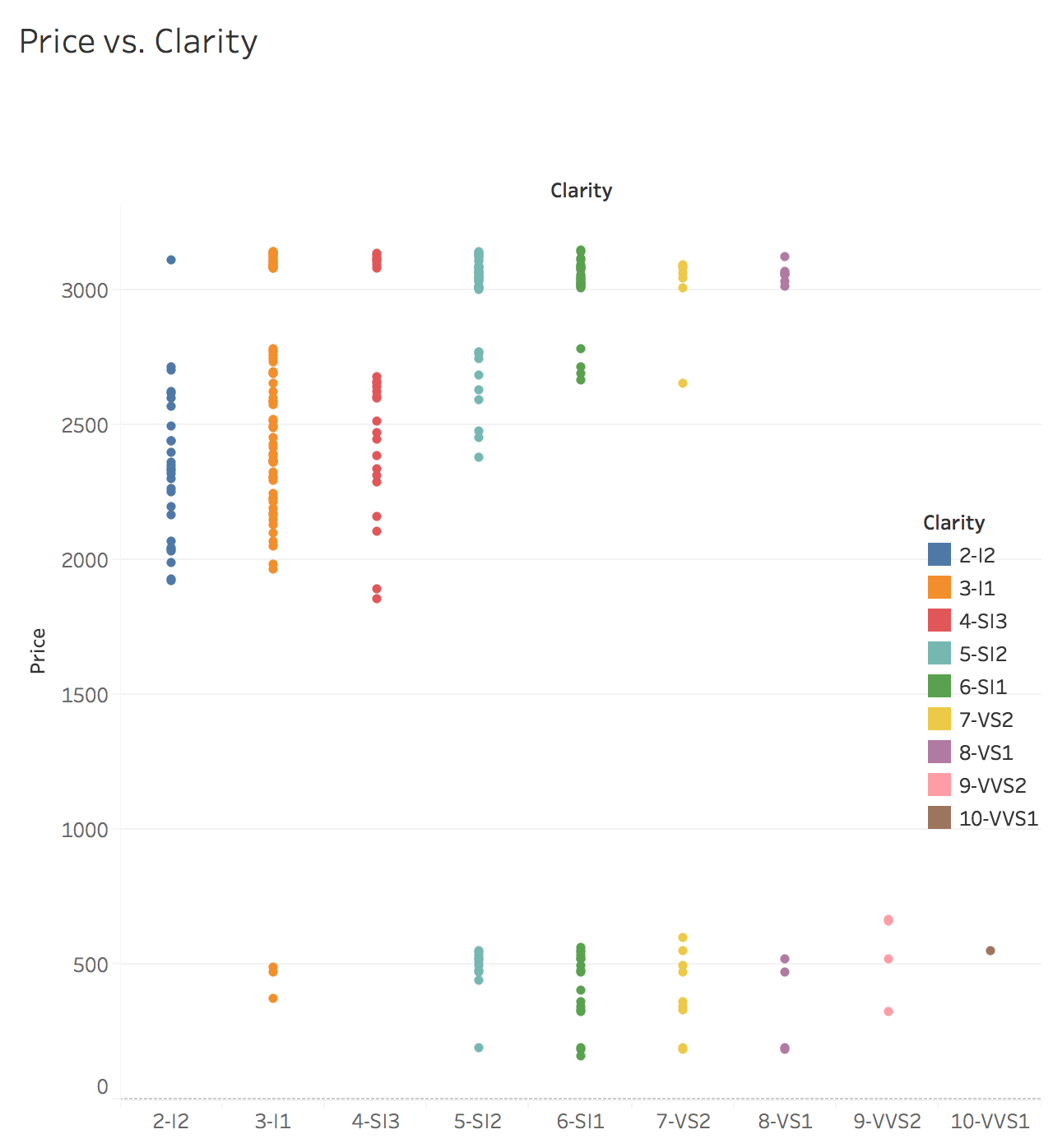


Cut

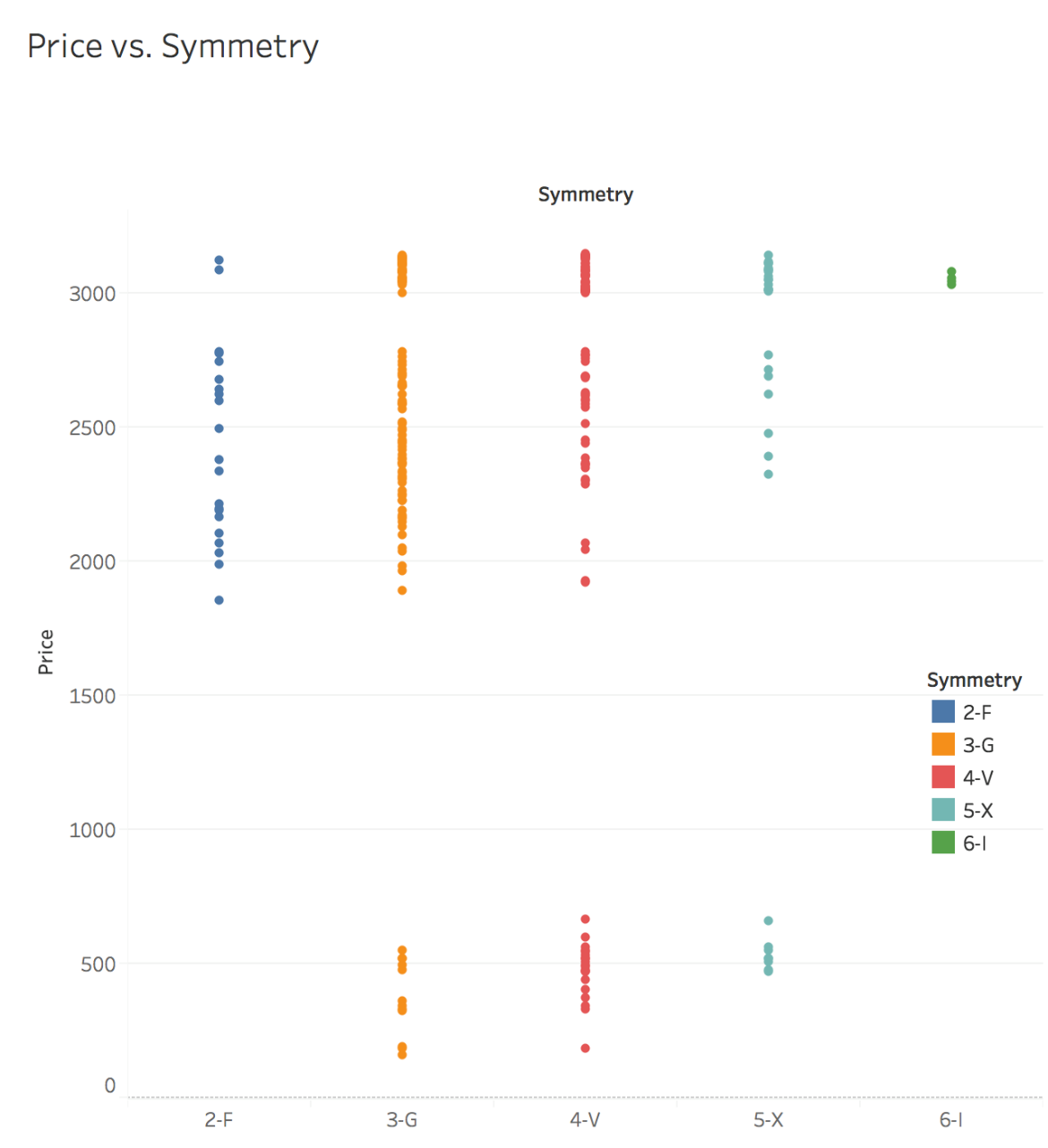


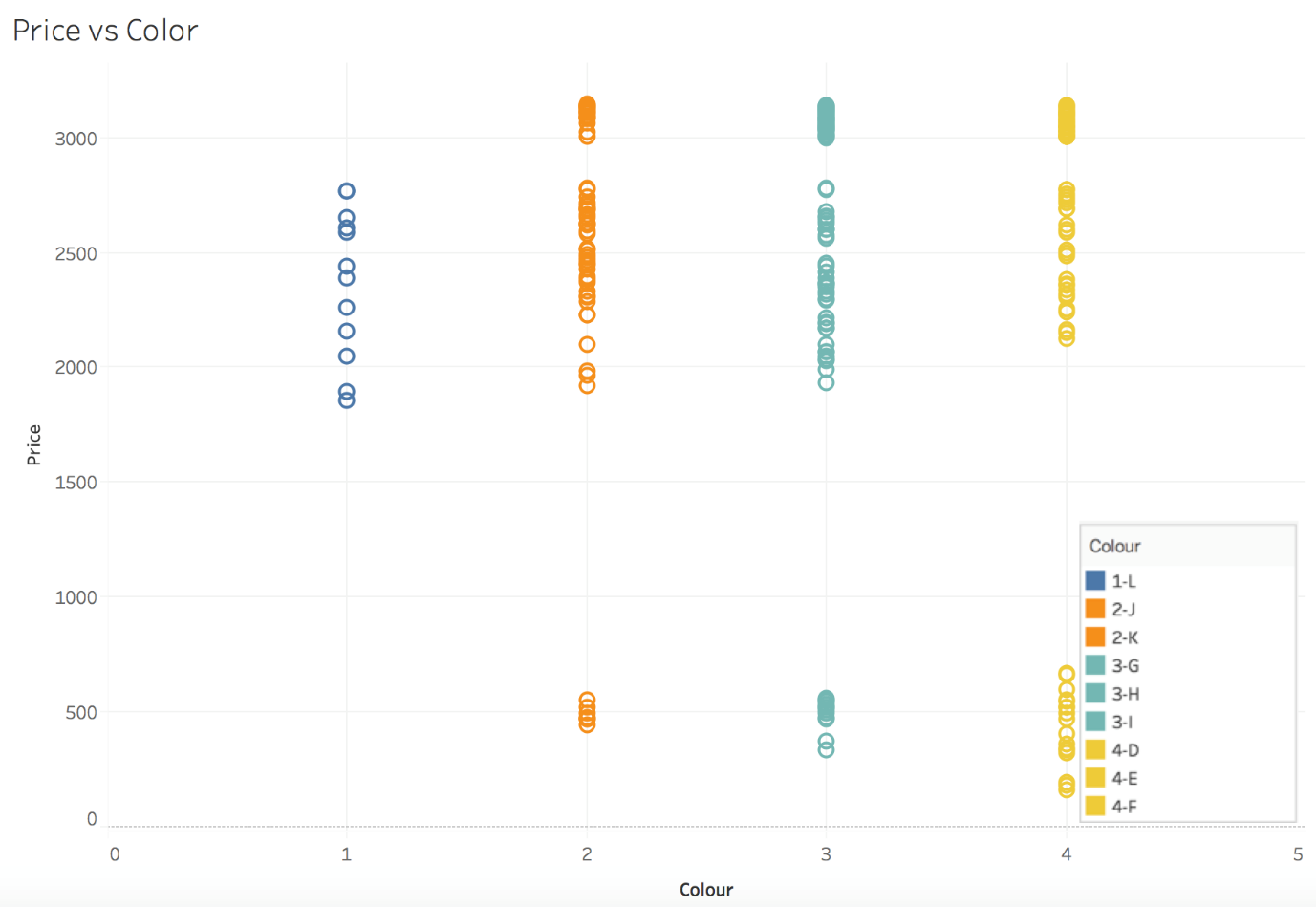
**Bivariate analysis**

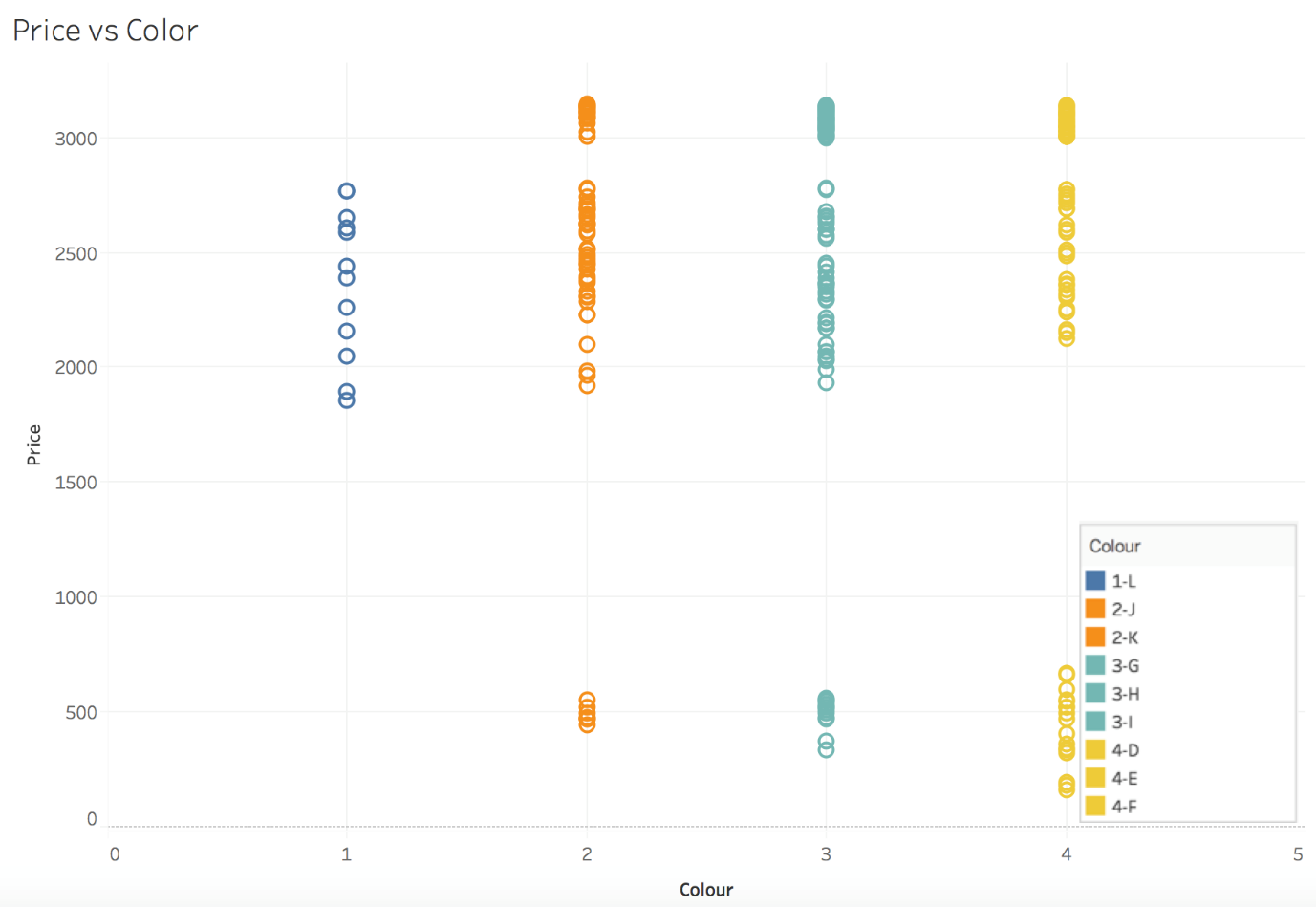












**Multi-collinearity**

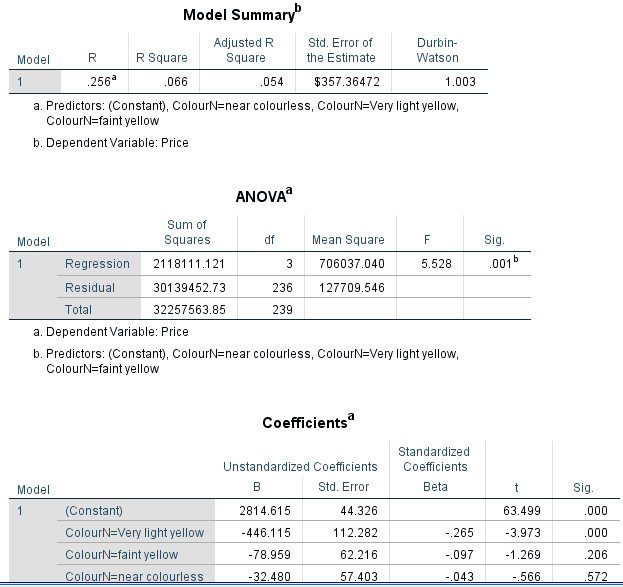
Take screenshot from PPT

**Feature Engineering**

**Start with Cluster formation Slide ..why we are concentrating on that cluster for further analysis**

Making every variable significant

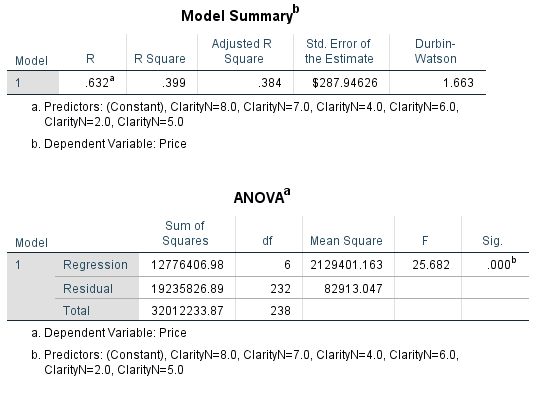
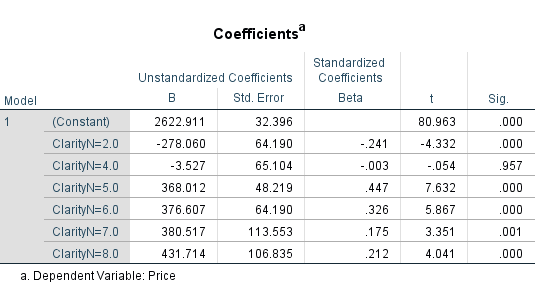
**Colour Intial regression result without categorising**



**After categorising**

REFER PPT

**Clarity**

Initial

Final

Refer PPT

**CUT**

Initial

A close up of a receipt

Description automatically generated

Final

REFER PPT

**Polish**

Initial

A screenshot of a cell phone

Description automatically generated

Final

REFER PPT

Symmetry

Initial

A screenshot of a cell phone

Description automatically generated

**Certification**

REFER PPT

A screenshot of a cell phone

Description automatically generated

Final

Refer PPt

**INITIAL MODEL and final model(describe why we removed symmetry – multicollinearity )**

**Refer ppt**

**Conclusion & results – Refer ppt**

**Residual plots – I don’t feel it is neccesssay , If u want generate it in spss and attach**