3.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 |

We started by checking all the independent variables whether the levels are significant for determining the price. As all the variables p-value is less than 0.05, we concluded that all the variables are significant enough to be used in the model

* MULTICOLLINEARITY



Firstly, categorical variables are converted into type numeric for checking the correlation. Then the collinearity plot is created. Since, this practice is not truly reliable for categorical variables, we decide to consider it as an indicator and do not totally rely on it. In the correlation plot the values are ranged from 1 to -1. Blue color states positive correlation and red color stated negative correlation

Bigger the circle more the correlation between those variables

Carat and Clarity: Clarity of diamond will decrease as per the increase of carat in diamond as both the variables are inversely proportional and it shows negative correlation.

Polish and Symmetry: The more you polish the diamond it would be more symmetrical and same applies with *Cut and Symmetry* and so it’s showing strong positive correlation between them