**MBA6636**

**Assignment 01**

**Priyal Chandrakant Jawale**

**Student Number - 3669751**

**Date: 09/25/2021**

**Summary of findings**

**Categorical variables**

1. Cut

The bar plot illustrates that the “ideal” cut diamonds are the most popular choice amongst the consumers. Whereas the “fair” cut diamonds appears to be least popular choice.

Chart, bar chart

Description automatically generated

1. Color

The bar plot which shows the colour of the diamond affecting the choice of consumer. The “I”, “J” and “D” color diamonds are significantly less in demand. It appears that “E”, “F” and “G” are high in demand and are closely distributed.

Chart, bar chart

Description automatically generated

1. Clarity

The graph depicts the clarity of diamond affecting the consumer buying the diamond. It is observed that as the quality of diamond increases, the number of consumers buying it decreases significantly. Hence, the overall demand decreases as the price goes higher.Chart, bar chart

Description automatically generated

**Continuous Variable**

1. Price

Plotting the dependent variable - Price.

Appears to be exponential relationship based on price of the diamond.

Shape

Description automatically generated with medium confidence

1. Carat

Strange visualization pattern, but it’s not flat can be used to visualize interesting correlations.

Chart, scatter chart

Description automatically generated

1. Depth

The scatter plot appears to be linear, wouldn’t give any interesting correlation results.

Chart, scatter chart

Description automatically generated

1. Table

The scatterplot illustrates a straight line which is basically flat. Thus, not to be considered for correlations.

Chart

Description automatically generated

**Simple histogram of diamond prices**

The graph illustrates that the concentration of observations decreases above $5000 threshold, this is a long tail distribution.

Chart, histogram

Description automatically generated

**Diamond Price Distribution by Cut**

The histogram shows that majority of the data appears to be distributed across diamonds in the Premium and Ideal categories. This appears to be the most popular categories among diamond purchasers for the Cut feature.

Diagram

Description automatically generated

**Price per Carat of different Cuts**

The graph shows how cut quality affects pricing per carat, it increases.

A picture containing text, map, cat

Description automatically generated

**Diamond Price Distribution by Color**

Diagram

Description automatically generated

**Diamond Price Distribution by Clarity**

Diagram, engineering drawing

Description automatically generated

**Diamond Price Distribution by Carat**

The scatterplot indicates that the price and carat weight have a positive, linear characteristics.

Chart, scatter chart

Description automatically generated

**Linear Price and Carat Relationship**

Chart, scatter chart

Description automatically generated

Relationship between continuous variable - Price and categorical variables - Cut, Clarity and Color

**Diamond Price according to Cut**

The conclusion drawn from the box plot is difficult, it appears that all the cuts have different costs. It doesn’t indicate how good or costly the diamond is.

Chart, box and whisker chart

Description automatically generated

**Diamond Price according to Clarity**

The results are significant, obtained outliers. Clarity appears to be meaningful variable for the model.

Chart, bar chart, box and whisker chart

Description automatically generated

**Diamond Price according to Color**

The graph illustrates that color appears to have an impact on the quality of a diamond and whether it will be costly. Color seems to be a meaningful variable.

Chart, bar chart, box and whisker chart

Description automatically generated

**Diamond Price per Carat according to Color**

The graph shows that the lower-quality diamonds seem to be more expensive. However, carat weight plays a major role in determining the price of the diamond.

Chart, box and whisker chart

Description automatically generated

**Correlation**

Price and Carat

As expected, the correlation between carat and the price is strong.

Chart

Description automatically generated

**Transformations**

Price – Log 10 Transformation

Chart, scatter chart

Description automatically generated

Carat – Log 10 Transformation

The scatterplot shows linearity.

Chart, scatter chart

Description automatically generated

Price vs. Carat and Clarity

The plot shows strong correlation between the variables. The higher the price, the better the clarity.

Chart, scatter chart

Description automatically generated

Price vs. Carat and Cut

The plot depicts that the Ideals dominate the cut, and there is a large variance under the category of cut.

Chart, scatter chart

Description automatically generated

Price vs. Carat and Color

From the given scatterplot, it appears that the variance in price distribution of color is small. Hence, the price and color have a strong correlation.

Chart, scatter chart

Description automatically generated

**Linear Regression Model**

It appears a good fit for simple linear model.

Chart, scatter chart

Description automatically generated

I believe this verifies the best line fit recommendations of variables, it appears that **Carat, Color, Cut** and **Clarity** are the most essential factors in determining the price of the diamond. This appears to confirm the use of the best fit model.