Project: Diamond Prices

Complete each section. When you are ready, save your file as a PDF document and submit it here: https://classroom.udacity.com/nanodegrees/nd008/parts/235a5408-0604-4871-8433-a6d670e37bbf/project#

Step 1: Understanding the Model

Answer the following questions:

Step 1 – Understand the data: There are two datasets.

1. diamonds.csv contains the data used to build the regression model.
2. new_diamonds.csv contains the data for the diamonds the company would like to purchase.

Creation of the multivariate linear regression model from diamonds.csv.

```
#Create a linear regression model for price with predictor variables carat, clarity_ord and Cut_ord
price_model <- lm (formula= price ~ carat+clarity_ord+cut_ord, data =diamond)

lm(formula = price ~ carat + clarity_ord + cut_ord, data = diamond)

Residuals:
    Min      1Q  Median      3Q     Max
-19245.7   -693.0   -105.3    542.7  10955.7

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -5255.223     30.320 -173.33   <2e-16 ***
carat        8363.417     13.565  616.55   <2e-16 ***
clarity_ord   457.802      3.901  117.37   <2e-16 ***
cut_ord       160.379      5.513   29.09   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1348 on 49996 degrees of freedom
Multiple R-squared:  0.8862,  Adjusted R-squared:  0.8862
F-statistic: 1.298e+05 on 3 and 49996 DF,  p-value: < 2.2e-16
```

So, the price = -5255.223 +8363.417*carat +457.802*clarity_ord +160.379*cut_ord

1. According to the model, if a diamond is 1 carat heavier than another with the same cut and clarity, how much more should I expect to pay? Why?

   If a diamond is 1 carat heavier than another with the same cut & clarity a price of 8363.417 extra must be paid as the coefficient of carat as per the linear model is 8363.417

2. If you were interested in a 1.5 carat diamond with a **Very Good** cut (represented by a 3 in the model) and a **VS2** clarity rating (represented by a 5 in the model), how much would the model predict you should pay for it?

   As per the linear regression model,
   Price = -5255.223 +8363.417*carat +457.802*clarity_ord +160.379*cut_ord
   = -5255.223 + 8363.417*1.5 +457.802*5 +160.379*3
   = 10,060.0495

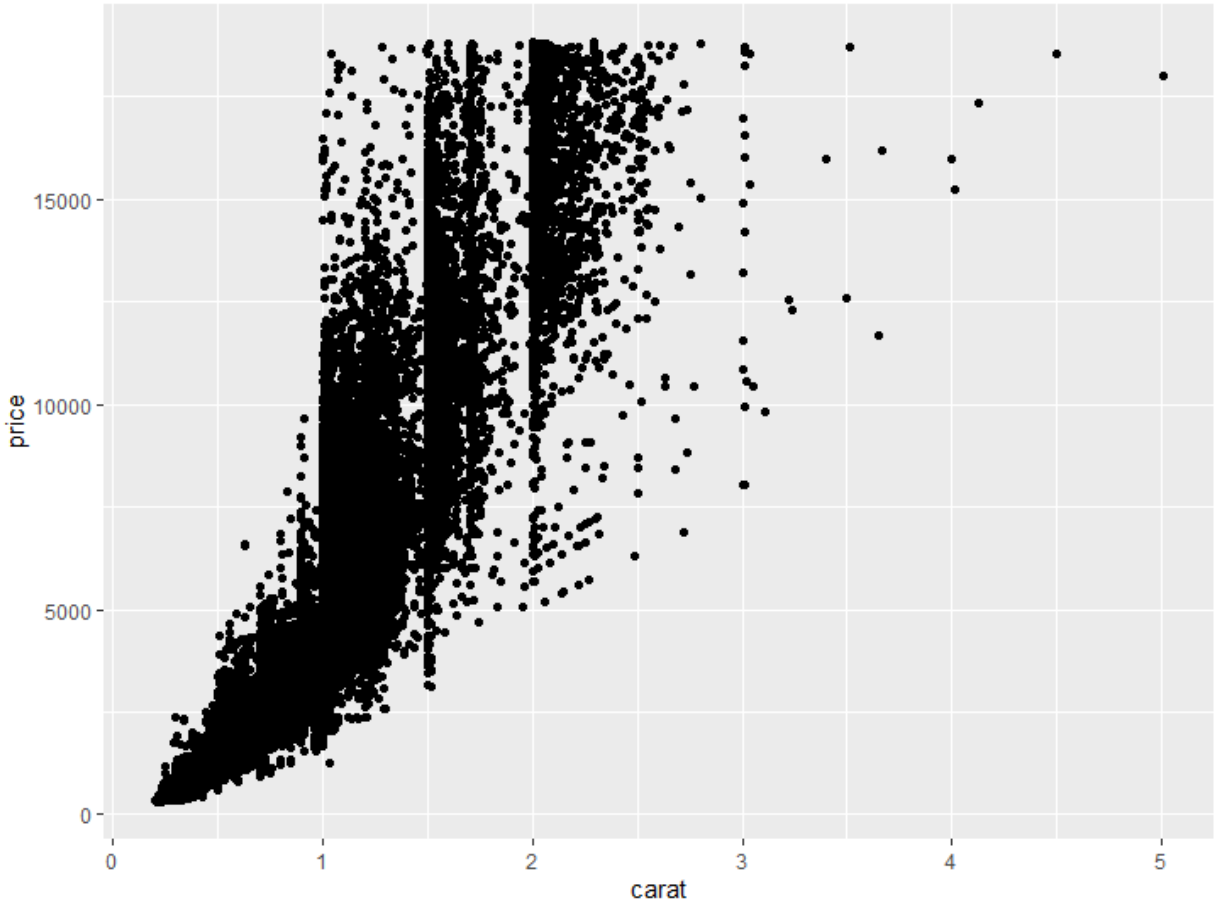
Step 2: Visualize the Data

Make sure to plot and include the visualizations in this report. For example, you can create graphs in Excel and copy and paste the graphs into this Word document.

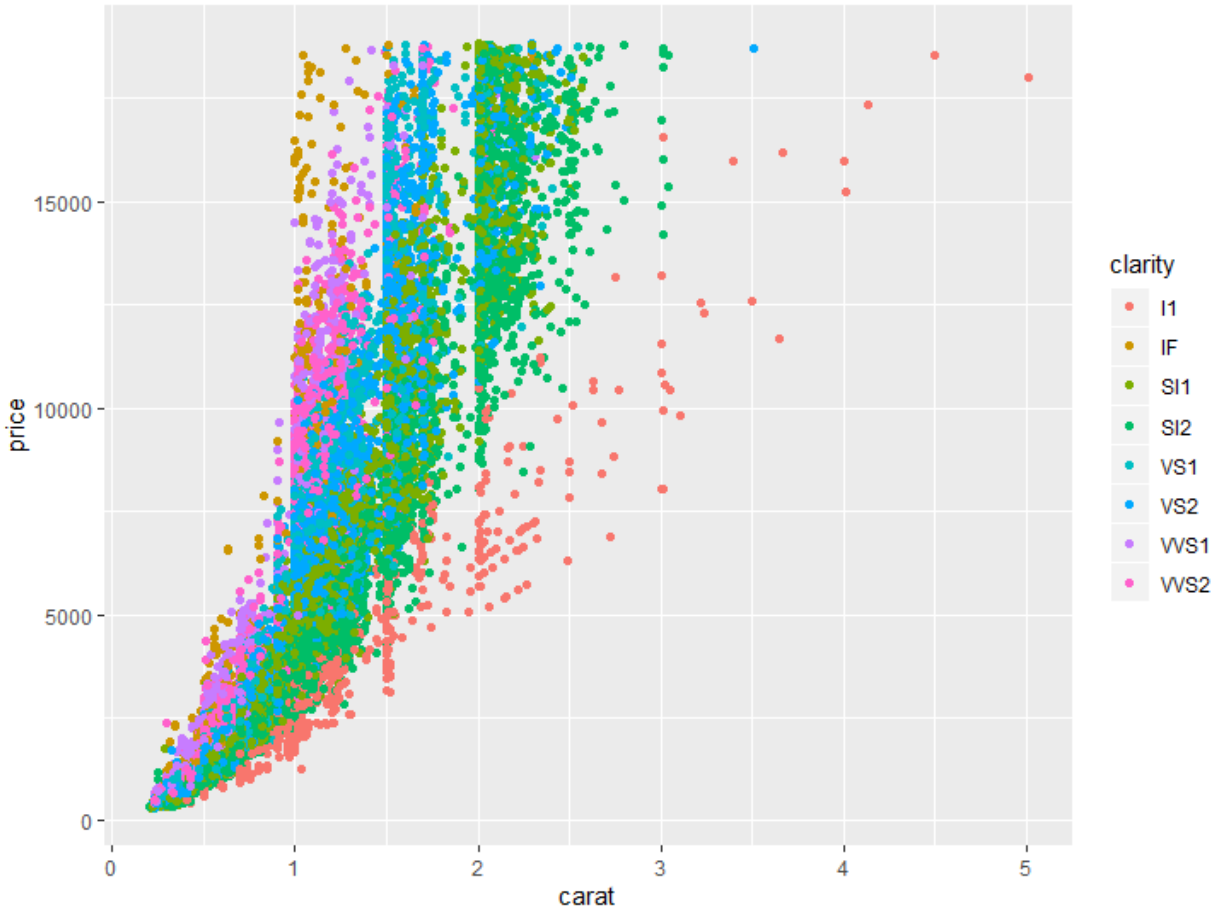
1. Plot 1 - Plot the data for the diamonds in the database, with carat on the x-axis and price on the y-axis.

   #Plot 1 - Plot the data for the diamonds in the database, with carat on the x-axis and price on the y-axis.

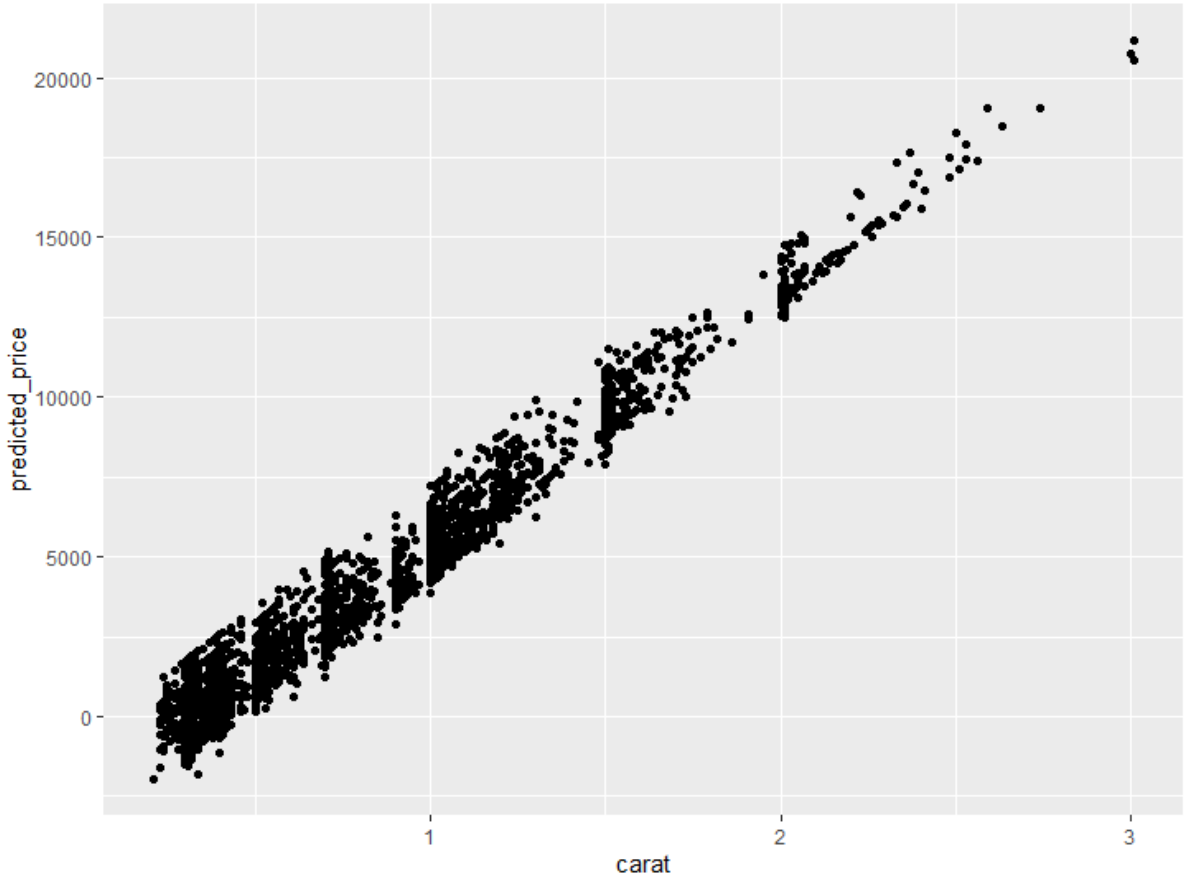
   ggplot(diamond, aes(x=carat, y= price))+ geom_point()



```
# see the color of points with respect to clarity
ggplot(diamond, aes(x=carat, y= price, color=clarity))+ geom_point()
```



2. Plot 2 - Plot the data for the diamonds for which you are predicting prices with carat on the x-axis and predicted price on the y-axis.
   #Plot 2 - Plot the data for the diamonds for which you are predicting prices with carat on the x-axis and predicted price on the y-axis.
   ggplot(diamond_price_pred, aes(x=carat, y= predicted_price))+geom_point()



   - **Note**: You can also plot both sets of data on the same chart in different colors.

3. What strikes you about this comparison? After seeing this plot, do you feel confident in the model's ability to predict prices?
   The linear regression is not an appropriate model as we have a mix or categorical and continuous variables. We have converted the categorical variables into ordinal variables and used them as continuous variables which is not a good practice without establishing the linear relationship in the data.
   Also if we observe the plot 1 closely, we can see some classification problems (grouped data) in this case we should have separated the class and applied regression model.
   Also the predicted model seems not very accurate, as significant predictions are negative values, not practical enough.

Step 3: Make a Recommendation

Answer the following questions:

1. What price do you recommend the jewelry company to bid? Please explain how you arrived at that number.

```
#bid based on predicted price.
sum <- sum(diamond_price_pred$predicted_price)
bid <-0.7*sum
```

Perform sum on all the predicted prices. (11730242)
Since this is the consumer price, take 70% of the sum and that is the amount a company should bid for diamonds. (8211169)