Extra Credit W2020 Due Friday March 6th @ 11:59 pm

Effect of Wine Critics' Ratings on Prices of Bordeaux Wines. In this assignment, we are interested in the effects of an American wine critic, Robert Parker and an English wine critic, Clive Coates on the London auction prices of Bordeaux wines from the 2000 vintage. The Wine data is posted on week 8. This data contains "Price" as a response variable and 7 predictors. The first two predictors are numeric, while the rest are categorical.

The variables in the data are as follows:
Y = Price = the price (in pounds sterling) of 12 bottles of wine
x1 = ParkerPoints = Robert Parker's rating of the wine (out of 100)
x2 = CoatesPoints = Clive Coates' rating of the wine (out of 20)
x3 = P95andAbove = 1 (0) if the wine scores 95 or above from Robert Parker (otherwise)
x4 = FirstGrowth = 1 (0) if the wine is a First Growth (otherwise)
x5 = CultWine = 1 (0) if the wine is a cult wine (otherwise)
x6 = Pomerol = 1 (0) if the wine is from Pomerol (otherwise)
x7 = VintageSuperstar = 1 (0) if the wine is a vintage superstar (otherwise)

- **Answer all the questions.**
- **This is not a group assignment.**
- **Help is not provided by TA nor by the professor.**
- **Create all plots based on ggpolot2 library.**
- **Create a PDF file with all the questions and their answers.**

This code helps creating an interaction plot using ggplot:
ggplot() +aes(x = 1st categorical variable, color = 2nd categorical variable, group =2nd categorical variable , y = Response) +stat_summary(fun.y = mean, geom = "point") +stat_summary(fun.y = mean, geom = "line")

Question (1):

A. First, using ifelse function, convert x3, x4, x5, x6 and x7 variables into categorical with "Yes" instead of 1 and "No" instead of 0. Name the new variables as P95andAboveNew, FirstGrowthNew, CultWineNew, PomerolNew and intageSuperstarNew respectively.
B. Create ggpairs plot for the response and the other two numerical predictors. What did you notice?
C. Create a MLR (call it m0) using the two numerical predictors only. Study the summary, anova, vif and diagnostics of the model.
D. Use leveragePlots and mmps on m0. What did you notice? Run powerTransform and inversResponsePlot functions on m0. What do you need to do to make m0 a better model? Do it.

E. A statistician suggested to use log transformation on the response variable and the two numerical predictors. Do you agree or disagree with that suggestion? Try it first then compare to m0.

Question (2):

A. Create a side by side box plots of the variable Price for each of the categorical variables created in Question 1 Part A. Which of these predictors you think are good predictor for the variable "Price"? Why?
B. Create a MLR (call it mcat) using all categorical predictor to predict "Price". Check summary, anova, vif and diagnostics of the model. Does the MLR summary agrees with your answer in part A?
C. Interpret the y-intercept and all the partial slopes in your mcat MLR.
D. Create 5C2 = 10 pairwise interaction plots. Which ones you think should be added to your model mcat?
E. Create a new MLR using the categorical predictors and the significant pairwise interactions (call it mcat2). Check summary, anova, and diagnostics of the model.
F. Conduct a Partial F-test between mcat and mcat2. What do you conclude?

Question (3):

A. Create a MLR (call it m1) using the suggested transformation on the numerical variables in the data along with the categorical predictors listed in your MLR mcat (No interaction terms). A total of 7 predictors. Check summary, anova, vif and diagnostics of the model.
B. Create another MLR (call it mfull) using the suggested transformation on the numerical variables in the data along with the categorical predictors listed in your MLR mcat with the significant interaction terms). A total of 7 predictors. Check summary, anova, and diagnostics of the model.
C. Interpret the y-intercept and all the partial slopes in your mfull MLR.
D. Which of the predictors need to be dropped from mfull?
E. Create a MLR (mred). Conduct partial F-test.

Question (4)

A. State your final MLR based on your answers to the previous three questions.
B. Interpret the y-intercept and all the partial slopes in your final MLR.
C. Identify the Unusually highly priced wines and the Unusually lowly priced wines based on your final model.
D. Identify the wines that can be considered good leverage points in your final MLR.