

HW 5 Winter 2020 Due Friday March 6th, 2020 @ 5 pm
Final Project Related Questions: Prelim Report (Just answer the questions). Individual submission.

Q1) Report the following from your training data used to create your latest MLR:

- a) State you name and your group name.
- b) The dimension of your training data after cleaning the NAs
- c) Summary statistics of your response variable only.
- d) How many predictors used to create your latest MLR.
- e) Classify your predictors: Categorical or Numerical: Template Table

No.	Categorical	No.	Numerical
1		1	
2		2	
3		3	

- f) Report your latest R^2 and latest Rank on Kaggle.
- g) Create matrix plot for your variables.
- h) Create corrplot of your numerical variables.

Q2) Have you used any transformation on your predictors or on your response variable?

- a) If yes, explain how did you decide what transformation to be used. List the variables and the transformation function used in your latest MLR. (Provide proofs of your work).
- b) If no, explain why the suggested transformations did not work out for your latest MLR. (Provide proofs of your work).

Q3) Report the following from your latest MLR:

- a) Anova table of your MLR
- b) Sort your predictors by their importance or controbutions
- c) Report R^2 and your R^2 -Adjusted of your MLR using the training data.
- d) Report the VIF of every predictor in your MLR make sure you have no multicollinearity violation (No predictor has a VIF exceeding five). Use the following template

No.	Variable Name	VIF
1		
2		
3		

Q4) Report the following:

- a) Diagnostics six plots of your latest MLR. Comment on how well or how bad your MLR.
- b) Identify your bad leverage points. How many and what are you planning to do to fix this problem.
- c) Identify all good leverage points based on your latest MLR. Any comments?
- d) Report your summary statistics of the predicted response variable in both training and testing data sets. Any comments?

Q5) Apply the step function and regsubsets function in r on your latest MLR and use it to answer the following: Need library “leaps”

- A. Identify the optimal model or models based on R^2_{adj} , AIC, BIC from the approach based on all possible subsets.
- B. Identify the optimal model or models based on AIC and BIC from the approach based on backward selection.
- C. Identify the optimal model or models based on AIC and BIC from the approach based on forward selection.
- D. Compare and contrast the models chosen in (A) (B) & (C). Check those which are similar and those which are different “maybe”.
- E. Recommend a final model. Give detailed reasons to support your choice on final model.
- F. Interpret the regression coefficients in the final model. Is it necessary to be cautious about taking these results too literally?