DATA 624 Spring 2019: Project-2

Project #2 (Team) Assignment and Scenario- these are the parameters

This is role playing. I am your new boss. I am in charge of production at ABC Beverage and you are a data scientist reporting to me. My leadership has told me that new regulations are requiring us to understand our manufacturing process, the predictive factors and be able to report to them our predictive model of PH.

Please use the historical data set I am providing. Build and report the factors in BOTH a technical and non-technical report. I like to use Word and Excel. Please provide your report in a Word readable format and your predictions in an Excel readable format.

I also rely on a colleague for advice. She is very data savvy and can provide info on good code form to me, and just make me feel better about a technical solution. Please provide all your code and technical dialogue so she can review it. She should be able to quickly cut and paste into R studio. NOTE, include R library calls in your code.

Questions? We can discuss. As always, I am busy (as your boss) so I really want you to take the ball and run with it the best you can. But, I will answer as I can. Let's talk more in our weekly meeting next Tuesday.

Should be a great project. Don't over think it, pretty straight forward based on what we have been learning.

You have MORE than ample time to complete, **5/20/19**. Bonus points for EARLY submissions.

Thanks for your dedicated efforts!

StudentData.xlsx - this is your modeling data

StudentEvaluation- TO PREDICT.xlsx - I want you to predict the PH on this set

Data Dictionary.xlsx - Not a full data dictionary, but variable names, data types, etc.

While prediction accuracy is very important. I really want to understand your technique. NOT JUST YOUR CODE! I want you to explain to me WHAT you are doing and WHY. Make it SIMPLE for me to read and comprehend. This is a report to your boss. IMPRESS him. Yes, you can add code, but this is secondary – your commentary of your process and the results are core.

Note to team leaders, you are responsible to turn in your team's project - late submissions will be heavily penalized.

Note to team leaders, you are responsible to turn in your team's project - late submissions will be heavily penalized.

Some Clarifying Points – What you will turn in

1) A Word readable file with a detailed explanation of the problem you are trying to solve, how you are approaching it, your results, interpretation of your results, etc. You have creative license here. Preferably you will also include an appendix with the R code I can cut and paste into R Studio and run. If you already have it embedded in your report and I can still copy/ paste that is fine, but it should be in a format (like Courier) where I can simply copy from the doc and paste into R Studio.

2) An Excel readable file with your predictions. Please put your predictions alongside the independent variables. This is important as there is no case/row identifier. Depending on how you handle outliers this could affect your row count. Also, it will allow me to see your imputation methods that you describe in your report if you chose to go that route.

3) Optional - R files (*.R) or Rmd

Final Assessment of Predictions

I use the average RMSE for your predictions to determine your overall accuracy. I will also look at the CV(RMSE) to determination the variability of the errors, where CV is the coefficient of variation.