

Predict 411, Unit 1
Homework Grade Sheet

Brent Young

<u>Section/Comments:</u>	<u>Points Allocated:</u>
Data Exploration (40 Points): Good	40
Data Preparation (40 Points): Good	40
Build Models (40 Points): Good	40
Select Models (40 Points): Good	40
Data Step (40 Points): Good	40
Scored File (50 Points):	46

One of the top 5 models for the class, great work.

Your model was ~24% away from my best model based on my combination of metrics. A maximum of 109.3, minimum of 34.4, and an average of 80 showed a good distribution.

I suspect the reason your model was so strong was because of your data preparation stage (decision trees, random forests, imputations, and handling of outliers – MICE package), your data exploration was also strong (correlations, histograms, boxplots, transformations) capturing the distribution. I also appreciate your exploration of multicollinearity. Also, comparing models using multiple metrics was a good idea.

Note, I only give the top one or two models in a class a 50/50 score, and then scale from there (and yes, I'm splitting hairs, but it makes it somewhat competitive and entertaining for the students at the top of the class).

Bonus: *(please note bonus on cover page)* 20 (per email)

General Deductions: N/A
(See general notes below.)

Total (250 Points): 266/250

General Notes:

- Note, some variables having counterintuitive coefficients is because it is in the presence of other variables. For example, because homeruns and triples are variables; doubles having a negative coefficient is plausible because it is *NOT* a homerun or a triple. Likewise, double plays indicates that runner were on base to begin with – which is not good either (from the defense's perspective). This is similar to suppressor variables (although not exactly the same).