

Rubric for Statistics 6372 Project
 Name of Students: Sam, Hari, Joe

Total Score: ___95___ out of 100 points

Following Guidelines (10 points)	
Was the final project turned in on time?	
	Total: ___10___ out of 10 points

Project set up and Data Collection and Cleaning	
(5 points) Were explanatory and response variables identified? Was the type of each variable mentioned? Was a reference given for the data set?	Yes.
(5 points) Was the data set described appropriately? If only part of the data were used, is the procedure for selecting that part described (Both variables or observations)? Was missing data discussed and explained how it was handled.	Yes.
	Total: ___10___ out of 10 points

Data Analysis and Model Building	
(10 points) Were appropriate descriptive statistics shown for various aspects of the data (response variables, explanatory variables, assessment of correlations, multicollinearity assessments? Were obvious issues discussed, mentioned, and/or attempted to address.	<p>Quick comment about transformations. When dealing with a skewed variable that has 0's. The transformation is commonly done on $\log(x+1)$. So just because you have 0's doesn't mean your lack the ability to do a log transform.</p> <p>In terms of the imputation, the flow of the analysis could be tweaked a little. Imputation would be a part of the data cleaning step to get an analyzable file. I would discuss the imputation properties and approach first, let people know you are going to be considering two versions of imputation, then move forward with the two data sets in tandem, summary statistics, discussing needed transformation, and assumptions. I understand that some of that is kind of hard to do sense you had to answer if the sequence of procedure matters. I would keep that up front with general discussion of imputation. With imputation behind you, you have the data set to move forward with EDA etc. It'll flow</p>

	<p>better I think.</p> <p>Summary statistics were provided and correlations</p>
<p>(10 points)</p> <p>Were appropriate graphical displays used for the techniques used? Was the display appropriate? Were any special or unusual features noted and explained?</p>	<p>Summarizing the missing through MI was nice addition. See my comments at the bottom for additional ideas.</p>
<p>(10 points)</p> <p>Discussion of feature selection approaches (both algorithm and common sense logic), model fitting criterion, or data reduction techniques. Interpretation of PCs if applicable.</p>	<p>PCs were provided as a secondary mentioned. I commend the effort. I would have swapped figure 18 with the prediction performance of your PCs. You could spare the details, but it would show the performance was not better rather than you just saying it. The scree plot is helpful to choose the PCs but if you're not going into any detail then the final prediction performance is probably what you want to show there.</p> <p>Again imputation look was good.</p>
	<p>Total: <u> 28 </u> out of 30 points</p>
Inference and/or Prediction assessment	
<p>(10 points) Was the appropriate inferential procedure used? Were the null and alternative hypotheses given if appropriate? Were the population and sample correctly identified? Were the results interpreted in context?</p>	<p>Cross validation procedures were done appropriately. Even though you compared a single vrs 25 imputed data sets, I would like to have seen an acknowledgement that the reason for comparing them in the first place is that performance using the 25 could be overly optimistic through CV.</p> <p>I would have like to seen a little more discussion on the choice of priors. Was the data set collected in such a way that the percentage breakdowns used are going to be realistic (aka is this a random sample)</p> <p>How robust was your prediction model if you changed the priors? You guys have decent balance with a lot of data so I don't think you really have a problem, but it could be an issue for other data sets.</p>
<p>(10 points) Were the assumptions checked with the appropriate statistics or graphics?</p>	<p>Assumptions were checked and an equal covariance matrix test was performed.</p>

(10 points) Was the appropriate output from software shown in order to support conclusions of the inferential procedure? Was cross validation used or validated on previous data set if prediction was the main goal.	CV tables were provided
	Total: __28__ out of 30 points

Conclusions (10 points)	
Were weaknesses in the study pointed out? Were the conclusions given in the context of the appropriate population from which the sample was drawn?	Yes.
	Total: __10__ out of 10 points

Clarity (10 points)	
(5 points) Is the report clear and concise? Did the report have sections? Did paragraph transitions flow smoothly from one paragraph to the next? Were words overused? Were there excessive spelling or typographical errors?	Overall, I thought the paper was well written and put together. My comments from before still stand. I think you could tweak the order just a little bit to smoothen things out.
(5 points) Did graphics have appropriate titles and labeling? Were figure captions appropriate?	Yes
	Total: __10__ out of 10 points

Additional Comments:

What I think would have been a good exercise to conduct given that you guys did 2 scenarios one imputation versus 25 already, go ahead and assess what the performance is like on the reduced “no missing” data set. Obviously its less but it would still be curious to see. Given your analysis what do you think would come from that? Maybe we can talk after class just to discuss not actually do.