Final Group Project Feedback

Group Number: 3	**Overall Grade: 80**
Team Members: Michelle Kim, Zhenyu Zhang, Georgina Jacko, Harvey Lee	

Model1: Logistic Regression	
Undertake EDA to understand data and select appropriate covariate sets (30%) **[26/ 30]**	A very thorough EDA and your insights regarding the data were well presented and discussed. I really liked your age distribution histograms, although these would be more relevant to present in relation to task 2 where we are particularly interested in Age. Possibly boxplots or a table summarising survival by covariates would be most relevant here. Your handling of missing data is well explained.
Constructs multivariable models and appraises model fit (40%) **[32/ 40]**	You explain and justify the variables for inclusion well and the first few steps are very clear and logical. I got a little bit lost at the point that you refit the data to the unimputed dataset. This was meant to come just at the end once you had decided on your final model. Your diagnostic checks were very thorough but it wasn't completely clear to me which model was being checked. You note some influential observations but then it wasn't clear from the report whether these were excluded or not from analyses. Overall based on both your rmd files I see good evidence of understanding of the modelling process and a thorough assessment of the relevant model metrics.
Interpretation (20%) **[16/ 20]**	You show very clear discussion of the model metrics and interpret them well. There were a few points that were a little inconsistent. Eg. sentence "The p-value of 0.3366 suggests that the increase in deviance (loss in fit) by removing these four predictors is statistically significant" It is not clear where the p value comes from and usually $p<0.05$ denotes statistical significance. It would also have been good to see a bit more discussion regarding the context and practical applications of the model. Also needed a bit more reference to some of the relevant literature. I note your references at the end, but if possible refer to these throughout (eg. by citing the author or number them) to support your discussion.
Presentation (10%) **[7/ 10]**	Overall this was a well presented analysis and it was clear to see what you had investigated. You could have been a little clearer regarding the definition of your final four models that were compared and it is helpful to label your figures with numbers and then refer to these in the text. Also you could work on presenting model results as a formatted table.
Overall Mark **[81/ 100]**	

Model 2: Survival Analysis	
Undertake EDA to understand data and select appropriate covariate sets (30%) **[24/ 30]**	A reasonable summary of the data and lovely KM plots showing survival time. A really useful one to present given the question was showing survival by a categorised version of age. This would also have given you confidence that it was appropriate to include age as a continuous variable – something that needed considering up front. As this was a slightly different question which focussed on age and survival, there was an opportunity to investigate the relationship between the other variables and age as well as considering the relationship with survival. (Your plot from part 1 did this!) The important covariates to include in your model are the ones that are related to both age and survival and particularly if there is supporting evidence of this from the literature.
Constructs multivariable models and appraises model fit (40%) **[31/ 40]**	Your modelling process was reasonable and you shows a good understanding of the models and were thorough in your investigation of the assumptions. As this task focussed on specific questions of interest it would have been good to focus your presentation of analyses a bit more toward this. The two key models of interest to answer the question were the univariable model with just age and one final multivariable model that controlled for other confounding factors to see whether these other factors reduced (explained) the effect of age on survival.
Interpretation (20%) **[16/ 20]**	Your report included a lot of general discussion of what the model metrics mean, but you could have focussed your discussion towards the question a bit more. The key thing in examining the output from the model is to evaluate whether the model is fit for purpose. In this case we weren't predicting, we were trying to understand the key relationships – specifically for age. Important for the model to fit well, but the main results of interest are probably the Beta value for age and how this changes when we control for other variables.
Presentation (10%) **[6/ 10]**	Overall some very nicely presented plots and you had a nicely formatted table in this section presenting model results. You could definitely cut back on some content though and just focus on the key findings that will answer the question.
Overall Mark **[79/ 100]**	