|  |  |
| --- | --- |
| Date/Time | 30st Mar 2017, 4:00pm to 5:00pm |
| Venue | SMU SISS Meeting Room 4.7 |
| Attendees | Prof Kam, Albert, Jun Liang, Russell |
| Agenda | 1. Share and get feedback on regression models to be used for report 2. Share and get feedback on R Shiny dashboard |

Consultation Meeting Minutes:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Task/Description | Person in Charge | Due Date |
| 1. | Feedback on regression models  Professor Kam mentioned that when using stepwise regressions, if variable has many categories, it is not necessary to create dummy variables because JMP Pro will already split accordingly.  Professor Kam also advised not to use “TSL Crew” as a bin for those videos with more than 3 actors because then it would be pointless to have listed out the actors for all the other videos. If the bin is used, it would be sufficient to get just the count of actors. The team would have to manually review the relevant videos again.  Professor Kam advised to remove videos with abnormally long length as these would represent outliers within the context of our sponsor.  Professor Kam advised the use of recursive partitioning for the variable measuring time of day. This is to bin time of day into several intervals to reduce the number of timings. He also demonstrated recursive partitioning as implemented by JMP Pro 13.  Regarding model fitting, Professor Kam commented the following:   1. First round of fit should just include the continuous variables. See the results from the Least Squares regression, then include all the categorical variables 2. With all the categorical variables included, run the stepwise regression – define the stopping rules and decide the conditions; use 95% confidence level 3. Note that VIF is not for categorical variables   Professor Kam commented that there may be too many categorical data, so it would be good to use dummy variables so that those with lesser effect can be excluded from the regression. Also, we should not exclude the ‘others’ category as there are too many instances and we would lose substantial data.  Professor Kam further suggested that we should review how academic papers do their regression analysis explanation to improve on ours, as well as to better improve our understanding on application of regression. | All | 3 Apr 2017 |
| 2. | Feedback on R Shiny dashboard  Professor Kam commented that there should be properly labelled headings for each column for better understanding.  Professor Kam commented that it is not important whether the dashboard is uploaded onto shinyapps or not because deployment is not necessary for the purposes of this module.  Professor Kam noted that there is a need to explain to people how to make use of the dashboard. For example, where there is an outlier, results may be affected drastically, and this is where Sparklines will come in useful. | All | 3 Apr 2017 |