IST 3420 – Introduction to Data Science and Management, Fall 2017, Chen

**Project Evaluation Form – Milestone 3: Data Analysis II**

**Submission Due December 1, 11:59 PM**

**Instruction:**

1. Collect and cleanse data. The project report should include:

* Introduction (refined from M2)
* Data Source and Collection (refined from M2)
* Data manipulation (refined from M2)
* Data summarization and visualization (refined from M2)
* Data exploration (newly developed)
* Predictive modeling (newly developed)
* Summarize Findings (newly developed)

Read the evaluation criteria carefully on the next page for the detail.

1. Use R Markdown to write your project report. You need to use proper Markdown syntax to format your report. Do not use MS Word or other format.
2. Please submit the following documents into Canvas:

* The project report written in R Markdown;
* The Word/HTML/PDF report that is directly generated from your R Markdown file;
* The Evaluation form with full project team information (see below table).

**Project Team Information (filled in by students)**

|  |  |  |
| --- | --- | --- |
| **Member name** | **Percent contribution** | **Activities completed by the member** |
| Adam Forestier | 30 % | Added visualizations, added predictive models, provided some analysis, met with professor, and attended team meeting |
| Brandon Jones | 30% | Provided analysis, attended team meeting, assisted with debugging, and met with professor |
| Nathaniel Williams | 30% | Provided analysis, added visualizations, added predictive model, attended team meeting, and met with professor |
| Parika Gupta | 10% | Provided feedback on work |

**Evaluation Summary – M3 (filled in by instructor)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Criteria** | **Target %** | **Comments** | **Evaluation** |
| * Refine your project report by addressing all issues on M2 commented by the instructor. * Submit all required documents: * The project report written in R Markdown; * The Word/HTML/PDF report that is directly generated from your R Markdown file; * The Evaluation form with full project team information. | 15 |  | 15 |
| * Correctly use various data exploration methods to explore your dataset. * Correctly use regression analysis methods to address some research questions and/or to get some solid conclusions that improve our current understanding of your dataset. | 20 | Figure “Average IMDb Score by Country”: better to use darker color to represent large values.  Correlation plot: it is not surprising to get a strong negative correlation between profit and budget, because profit = revenue – budget (or cost).  Need to interpret your regression result. What conclusions you can get from the regression? Do not simply give a brief summary in the last section. | 19 |
| * Properly use predictive analytics to solve some classification and prediction problems on your dataset. * The predictive analytics should include model building and tuning to get a “best” predictive model for your dataset. * At least use 4 prediction and classification models, properly visualize these models and compare their performance. | 40 | Why did you “Use simple splitting to predict imdb\_score”?  Compare your predictive models in the predictive modeling section, not the section that summarizes your whole data analysis. | 39 |
| * Summarize the key findings and provide suggestion for business improvement. * Explain implications or benefits your project can bring to the stakeholders. | 15 | Need to have more contents regarding what contributions your project can provide. | 14 |
| * Format your project report in a professional way. * Write your project report by using appropriate Markdown syntax. * Your report tells an interesting story. * The whole document is well written. | 10 |  | 10 |
| **Total** | **100** |  | **97** |