Stats 131 Final Project Grading Rubric

When grading written reports, final projects, and presentations, there is always a subjective component to the grading. I have tried my best to clearly articulate my expectations and grading criteria with the following rubric.

The point deductions in brackets are **guidelines** that I will use in grading. For example, a description in the 'Basic' column says [minus ~10 pts]. A group could receive anywhere from minus 5 points (if the report is between Basic and Excellent) or as much as minus 15 points (if the report is between Basic and Needs Improvement).
In the 'Good' column, [minus ~0] means anywhere from 0 to 3 points.

The final project is 300 points total.

	Good:	Basic:	Needs Improvement:
EDA: Writing	Explanations are correct, complete, and convincing. Assumptions are made explicit and given justification. [minus ~0 pts]	Explanations are partially correct but incomplete or unconvincing. Assumptions are made explicit but not justified. [minus ~10 pts]	Explanations are illogical, incorrect, or incoherent. Assumptions are not made explicit. [minus ~30 pts]
Selection of data	Data selection has been approved by professor. At least one team member is interested in and knowledgeable about the subject and are aware of important subtleties of the data. [minus ~0 pts]	Data selection has been approved by professor, but team members are not interested in or knowledgeable about the subject. [minus ~5 pts]	The data set is different from the one discussed with the professor. The data set was never discussed or approved by the professor. [minus ~30 pts]
Background Information	Complete background information relevant for understanding the data is provided. All variables have been described. After reviewing the provided background information, a person not familiar with the subject will be able to understand the significance of each variable in the data. Unnecessary or irrelevant information is not included. [minus ~0 pts]	Background information relevant for understanding the data is provided. All variables have been described. Important contextual information is missing, for example a person unfamiliar with the subject would not know whether values are unusually large or small. Includes information irrelevant to understanding the data. [minus ~10 pts]	Background information is incomplete or missing significant details. Origin of data (who collected it, how it was gathered, when it was gathered) is incomplete Data is artificial or simulated. Variables are not defined or explained. [minus ~40 pts]
Exploratory Data Analysis: Motivation, Relevance	Analysis follows a theme and is motivated and guided by research questions. Analysis is appropriate to questions, is complete, relevant and informative.	Analysis is motivated by research questions. Analysis is appropriate to the questions but perhaps incomplete, not informative, or irrelevant to the research questions.	Analysis has no clear direction, motivation, or theme. Analysis is overly simplistic or irrelevant. Analysis lacks coherence and sections performed by different students can be easily differentiated.

	An interesting finding in one table or graph leads to follow-up questions and analyses. [minus ~0 pts]	An interesting finding in one table or graph is given a remark, but no further analysis is performed. [minus ~10 pts]	[minus ~30 pts]
EDA: Exploration of individual variables	Analysis of individual variables is appropriate and informative. Interesting and/or unusual features or outliers are explored and given appropriate commentary or treatment (e.g. removed from analysis if appropriate, analyzed separately, no special treatment was deemed necessary, etc.). [minus ~0 pts]	Analysis of individual variables is appropriate and informative. Interesting and/or unusual features or outliers are explored but not given an appropriate treatment. [minus ~10 pts]	Analysis of individual variables is inappropriate (e.g. treating a categorical variable as numeric) Interesting and/or unusual features or outliers are not explored. [minus ~20 pts]
EDA: Exploration of Relationships between variables	Relationships between variables are explored and explained and the analysis has depth. If appropriate, relationships between variables are explored and explained separately for different categorical groups within the data (e.g. the relationship between GPA and GRE score for History majors, vs the relationship between GPA and GRE score for Physics majors, vs Psychology Majors, etc) [minus ~0 pts]	Relationships between variables are explored and explained but the analysis is simplistic and lacks depth. [minus ~15 pts]	Relationships between variables are not explored. [minus ~30 pts]
EDA: Visual presentation	Plots convey information correctly with adequate and appropriate reference information. Features of plots are fully explained. (e.g. a correlation heatmap between a subset of 4 variables with explanation that reveal how X1 and X2 are related positively to each other while negatively with X3 and X4, and how it is relevant to the analysis) [minus ~0 pts]	Plots convey information correctly but lack context for interpretation or are cluttered with unnecessary information. Explanations are provided for each plot. (e.g. a correlation heatmap of all variables, pointing out the important relationships between X1, X2, X3, and X4, but includes too much irrelevant info by including variables X5-X10.) [minus ~10 pts]	Inappropriate choice of plots. Plots are poorly labeled. Includes irrelevant plots that do not convey meaningful information with little explanation. (e.g. a correlation heatmap of all variables without any explanation) [minus ~20 pts]
Data Modeling: the model	The model is relevant to the theme and research questions. An appropriate response variable is selected. Appropriate variables with potential relationships identified	The model is relevant to the theme and research questions. An appropriate response variable is selected, but the criteria for excellent is not met. [minus ~10 pts]	The model is not relevant to the theme or research questions. A model is fit without justification or explanation. [minus ~30 pts]

	in the exploratory analysis are used and included in the model with explanation. If necessary, new variables or features are created with explanation. [minus ~0 pts]		
Data Modeling: Prediction (if groups choose prediction)	An appropriate model is fit to the data and used to predict values. The prediction performance is evaluated correctly. Potential improvements are identified and discussed and acted upon if appropriate. [minus ~0 pts]	An appropriate model is fit to the data and used to predict values. The prediction performance is evaluated but may be incorrect. Potential improvements are not discussed or an obvious improvement is overlooked. [minus ~10 pts]	An inappropriate or questionable model is used to fit the data (e.g. logistic regression for a variable with more than 2 categories.) Prediction performance is not evaluated. [minus ~30 pts]
Data Modeling: Insight (if groups choose insight)	An appropriate model is fit to the data and used to gain insight on the relationships between variables. The assumptions regarding the suitability of the model are checked, evaluated, and discussed correctly (e.g. constant variance for a linear regression model). The insight (or lack of insight) from the use of the model is explained correctly and clearly. [minus ~0 pts]	An appropriate model is fit to the data and used to gain insight on the relationships between variables. The assumptions regarding the suitability of the model are checked but has mistakes. The insight from the use of the model is explained but has mistakes. [minus ~10 pts]	An inappropriate model is used to fit the data. [minus ~30 pts]
Data Modeling: Model Selection	Different potential models are considered and compared. The merits of each model are discussed. [minus ~0 pts]	Different potential models are fit to the data. The discussion between models has mistakes. [minus ~10 pts]	Only one model is fit to the data. [minus ~20 pts]
Presentation	Presenters are poised and well-spoken. Presentation summarizes key findings and insight. Length of presentation is between 5-7 minutes. [minus ~0 pts]	Presenters do fine, but lack polish. Presentation summarizes findings from report. [minus ~10 pts]	Presentation is extremely short (< 3 min) or too long (> 10). It is clear presenters have not rehearsed the presentation. Slides are difficult to read. [minus ~20 pts]