	Excellent	**Good**	**Fair**	**Poor**
Dataset Origins (4 pts)	Origins fully explained, including when and where the dataset is found. Some information describing why we might care and how that informs the work in the report included.	Origins explained in less detail, includes some information describing why we might care.	Origins briefly described, does not inform why we might care.	Origins missing.
"Lit Review" (4 pts)	Other work on the dataset is fully described, including concise yet thorough summaries of findings.	Other work on the dataset is described, summaries of findings too brief to be considered useful.	Other work on the dataset is loosely described, missing data analysis methods and summaries.	Lit review missing.
Citation (2 pts)	Dataset name, author, revision number, and URL described in sufficient detail.	Missing details, but dataset can still be found.	URL only.	Citation missing.
Risk Taking (10 pts)[1]	Actively seeks out and follows through on untested and potentially risky approaches (e.g., trains and evaluates neural networks, uses other techniques not discussed in class).	Incorporates new approaches (e.g., attempts to train neural networks).	Considers new approaches without going beyond the guidelines of the assignment (e.g., discusses how they might use other methods not discussed in class).	Stays strictly within the guidelines of the assignment (earns at least 4 points out of 10).
Innovative Thinking (10 pts)[1]	Extends a novel or unique idea or question to create new knowledge or knowledge that crosses boundaries (e.g., creates a dataset by synthesizing from multiple sources or webscraping).	Creates a novel or unique idea or question (e.g., discusses how one might include new data or alternative data sources into the data analysis).	Experiments with creating a novel or unique idea or question (e.g., mentions that alternative data could be used but does not go into detail).	Reformulates a collection of available ideas (earns at least 4 points out of 10).

[0]Students will only be scored on one of the Risk Taking or Innovative Thinking dimensions. Category descriptions taken from the Association of American Colleges and Universities' *Creating Thinking Value Rubric*.

[0]Students will only be scored on one of the Risk Taking or Innovative Thinking dimensions. Category descriptions taken from the Association of American Colleges and Universities' *Creating Thinking Value Rubric*.

Data Modeling (10 pts)	Methods considered appropriate for solving the problem at hand (e.g., logistic regression for binary classification) have been correctly applied to a cleaned, pre-processed (missing data, outliers accounted for) dataset.	Methods considered appropriate for solving the problem at hand have been correctly applied to a cleaned, pre-processed dataset. Methods may not be the MOST appropriate, but might work well for some problems.	Methods have been carefully applied to a dataset. Methods are not appropriate, but do work for some problems (e.g., linear regression applied to time series data or with a 0/1 outcome variable).	Inappropriate methods have been incorrectly applied to a dataset.
Model Evaluation (10 pts)	Appropriate measures (e.g., cross-validated test set error for linear regression, AUC for binary classification) of model fit have been applied and correctly used to select a final model.	Appropriate measures have been used to describe model fit, but are not linked with final model selection.	Inappropriate measures of model fit have been applied.	Model fit has not been evaluated.
Model Discussion (10 pts)	Results from the model have been clearly, correctly, and concisely stated (e.g., a list of significant predictors). Conclusions are made based on results, but are not overstated.	Results from the model have been clearly and correctly stated. There may be some extraneous discussion or overstated conclusions.	Results from the model have been stated, but it is difficult to understand them (including difficult to find them or having results stated out of order).	Model has not been discussed.
Data Modeling Description (10 pts)	All methods (both for data cleaning/wrangling and analysis) have been described and justified so analysis can be independently replicated.	Most methods have been described and justified so that others could independently find results consistent with the analysis. Mainly missing data cleaning processes.	Methods have been described, but the entire process cannot be replicated.	Data cleaning/wrangling and analysis methods have not been described at all.
Data Visualization (10 pts)	Report includes appropriate and informative data visualizations (including both plots and tables). Visualizations are free of correctable flaws with appropriate labels, font size, color schemes, must not be grainy or illegible.	Data and model are summarized with mostly appropriate and informative data visualizations, there are a few easily addressed flaws.	Data and model are summarized with a few appropriate data visualizations, but some are inappropriate.	Data and model are summarized with mostly inappropriate data visualizations.