	To earn 5 points for a category, your project **must** meet all criteria listed.					Points Earned	Comments
	10	**9**	**5**	**3**	**1**		
Code	Uses the *tidyverse* and *tidymodels* for fitting models, etc. includes comments as necessary. **Uses code folding** so huge blocks of code aren't displayed when knitting.				Multiple problems. Minimal or no use of *tidyverse* and *tidymodels*. No code folding.		
Narration	Includes introduction and conclusion sections. Good flow of text and narration throughout; at least some description/explanation for every piece of code/results. Easily readable. Makes an effort to explain/make sense to general audience. **Written like a paper.** Divided into reasonable sections.				Minimal narration throughout. Missing introduction and/or conclusion section(s). Difficult to follow. No explanation of concepts/models.		
EDA	Conducts some EDA. Explores outcome variable distribution. Assesses missing data patterns. Creates at least 3-5 plots or tables to explore **relationships among variables**.				No EDA, or EDA that is missing a crucial aspect. For example, an EDA (usually) should not consist only of histograms of individual variables.		
Organization & Structure	Submits a **link to a GitHub repository** containing your project, with .Rmd and knitted .html or .pdf. Includes at least one subfolder for data (with subdirectories). **Includes codebook and data citation.**				Poor organization or no organization. No codebook and/or no mention of/link to data source. .Rmd, .html, or .R files are absent. Data file(s) are absent **without explanation** *(if they are too large to include or limited by confidentiality, simply make a note).*		
Feature Extraction	**Creates appropriate models.** Handles categorical predictors reasonably (with dummy/one-hot encoding, PCA, etc.). Creates interactions if necessary. Justifies any variable transformations.				Makes a mistake in model creation -- includes but not limited to including predictors that are generated from the outcome.		
Data Splitting	Divides data into training and test sets with a reasonable proportion. Uses stratified sampling. **Uses cross-validation to fold training set.** (Ideally, stratified CV with repeats.)				No use of cross-validation OR no initial split of the data. (Both are crucial.)		
Model Fitting	Fits **at least** 4 model classes (random forest, KNN, boosted tree, lasso or ridge regression or logistic regression, SVM, neural network). **Tunes models across resamples.** Fits optimal tuned model to training and test sets.				Fits only one model class OR **does not tune models** OR does not choose an optimal model after tuning OR does not fit optimal model to training and test sets and evaluate performance.		

Total:

0