Project Rubric

Project Specifications	Metric for success	Developing - 0		Accomplished - 1		Exemplary - 2		Notes
README.md	Student has a clear readme, highlighting important aspects of the project.	Student does not have a readme, or has one that is just a copy of the notebook.	☐	Student has a readme that is clear and well organized. It outlines their data (sources, anything quirky or hard to understand, etc.), their process, and their recommendations based on results.	☐	Student has a readme with a clear and well organized outlines, conclusion, and recommendation section. Visualizations are present. Language and markdown lend themselves to succinctness.	☐	
Data Collection	Chosen dataset was relevant to classification.	Dataset was not relevant to classification.	☐	Dataset was simple (such as a Kaggle dataset not supplemented with any other data), but relevant to classification. Has at least 1000 rows and 10 features.	☐	Dataset was originally constructed (webscraping/api calls made) and relevant to classification. Has at least 1000 rows and 10 features.	☐	
Exploratory Data Analysis	Student explores data using visualizations and descriptive statistics.	Inadequate visualizations and descriptive statistics, OR, did not interpret these appropriately.	☐	At least 2 well-constructed visualizations presented along with descriptive statistics; these are interpreted correctly. Student(s) can speak knowledgably about their data.	☐	Student created at least 2 well-constructed visualizations, discussed descriptive statistics, and is able to interpret and compare or link them together to give a cohesive introduction to the data.	☐	
Hypothesis Testing	Student used and interpreted statistical techniques (t-tests, confidence intervals, ANOVA, effect size, power analysis etc.) to explore relevant distinctions in the dataset.	Did not correctly interpret at least 2 statistical techniques.	☐	Used at least 2 statistical techniques with correct interpretation of results.	☐	Used 2 or more statistical techniques to generate insightful conclusions about the dataset.	☐	Note: These can (and should!) relate to your visualizations. Test the hypotheses you form during your EDA!
Model Fitting	Fit at least one model. Summarize model impact and meaning.	Attempted basic model fitting (or forgot to model fit). Incorrect application. Misinterpreted results.	☐	Correctly fit a single model. Correctly interpreted model results. Summarized model meaning & impact.	☐	Compared multiple models. Fit models outside of class materials. Detailed numerical and visual analysis of models.	☐	
Preprocessing/Feature Engineering	Student processed data and created features to increase the predictive capability of the models.	Very little to no preprocessing, no feature engineering	☐	Student preprocessed data and developed new features.	☐	Student preprocessed data and developed domain-motivated features based on their EDA that contributed positively to their model's performance.	☐	
Business/Social Case & Conclusions	Student makes a business/social recommendation driven by data analysis.	No clear business/social case or no conclusion is present.	☐	Appropriately challenging business/social case with conclusion present and includes a business recommendation motivated by data analysis.	☐	Appropriately challenging business/social case with conclusion present, includes a data-driven business recommendation, and future steps for further analysis/recommendations.	☐	
Quality of Code, Github	Student creates a notebook (or several) that is well documented, clean, easy to read and understand, and has a neat, organized GitHub repo.	Code is not in Github, or repo is used improperly (only one or two commits in total, does not show workflow). Code may not run, or is hard to read, or is not commented.	☐	Code is on Github with frequent commits (and relevant commit messages). Code is clean, organized, and well commented.	☐	Code is on Github with frequent commits (and relevant commit messages). Code is clean, organized, and well commented. Markdown cells add clarity and outline the process. Functions are utilized where appropriate to increase readability and reduce repetition.	☐	
Presentation	Student delivers a clear, organized, well-thought out presentation that can be understood by a non-technical audience.	Slides are unclear, disorganized. Visuals are not legible to audience. Slides may be too text-heavy. Student goes over time.	☐	Student delivers a presentation that is mostly organized and clear in the allotted time. Overall, presentation is understood by audience.	☐	Student is engaging and presents a well-organized, clear, legible presentation in the allotted time. Visuals add to the presentation and are well explained. The presentation tells a cohesive story.	☐	

| Score (1-5) | 1 | 0 | | 0 | | 0 | | |