	Exceeds	Meets	Approaches	Incomplete or unaddressed
Dataset novelty		**10 pts** — The information provided in the dataset is completely novel (i.e., there is no existing Kaggle dataset that has similar content and scope).	**8 pts** — The newly contributed dataset shares several observations/variables in common with existing Kaggle datasets.	**0 pts** — The dataset is an exact match for one or more existing Kaggle datasets.
Documentation and directory completeness		**15 pts** — The GitHub repository associated with your project is organized and populated with all requisite documentation files according to TIER Protocol 4.0. <u>Only the main branch will be assessed.</u>	**12 pts** — The GitHub repository associated with your project is mostly organized and complete, though may include superfluous files, incomplete metadata, or misplaced processing code. <u>Only the main branch will be assessed.</u>	**0 pts** — The GitHub repository associated with your project is largely unchanged from the template provided to you at the beginning of the project. <u>Only the main branch will be assessed.</u>
GitHub utilization	**5 pts** — You and your partner engaged with GitHub in a way that goes beyond the workflows and best practices introduced during <u>Lab #9</u>.	**4 pts** — You and your partner used GitHub effectively and in accordance with the workflows and best practices introduced during <u>Lab #9</u>.	**3 pts** — You and your partner used GitHub sparingly to collaborate on your project. The repository shows evidence of few branches, commits, and merges.	**0 pts** — You and your partner did not use GitHub at all to collaborate on your project.

Script organization and annotation	**20 pts** — All scripts are well-organized, using section headers to delineate discrete tasks, and feature thorough annotations which describe both the *what* and *why* of the code.	**16 pts** — All scripts are logically organized, though may lack headers or clear sub-sections. Code annotation is present and sufficient to understand function.	**12 pts** — Many or all scripts are difficult to read, lacking any sort of logical organization and very limited annotation.	**0 pts** — All scripts are completely incoherent.
Code sophistication and originality	**15 pts** —Original code reflects mastery of the functions and other programmatic constructs introduced in lecture. There are only one or two instances of code borrowed from the web or written by generative AI tools.	**12 pts** — Original code demonstrates a solid understanding of the functions and other programmatic constructs introduced in lecture. There are several unnecessary instances of non-original code borrowed from the web or written by generative AI tools.	**9 pts** — Original code demonstrates some understanding of the functions and other programmatic constructs introduced in lecture. Instances of unnecessary non-original code are common, and the scripts heavily rely on generative AI tools.	**0 pts** — Original code is basic, demonstrating no engagement with lecture content. Code shows compelling evidence of being almost entirely borrowed from the web or written by generative AI tools.
Code functionality	**30 pts** — The code accomplishes the goals of the project with impressive efficiency. The code throws no errors, and always uses programmatic solutions where appropriate.	**25 pts** — The code accomplishes the goals of the project with no errors and no manual work-arounds where a programmatic solution is accessible.	**20 pts** — The code mostly accomplishes the goals of the project, though throws a handful of errors and relies heavily on manually repeated code.	**0 pts** — The code does not function at all.
Dataset cleanliness		**10 pts** — The dataset produced is tidy, with standardized column names, cleaned values, and properly handled missing data.	**8 pts** — The dataset produced is mostly tidy, with syntactic column names. Most columns contain cleaned values and missing data is handled properly.	**0 pts** — The dataset produced is non-tidy in several ways, featuring unstandardized and non-syntactic column names. Most columns contain unclean values and missing data is not handled properly.

Kaggle page organization and completeness	**10 pts** — In addition to the requirements listed under **Meets** column, your Kaggle page must also: • Be well formatted and visually appealing • Specify an appropriate Creative Common license to let people know how they're able to use your data that is consistent with the licensing associated with your API and web scraping sources	**8 pts** — The public Kaggle page that you build for your dataset features each of the following: • A subtitle • A relevant image • A useful description • Metadata for each .csv file you upload • Information on provenance with link to GitHub repo • An example analysis notebook	**6 pts** — The public Kaggle page that you build for your dataset is lacking one or more of the requirements listed under the **Meets** column.	**0 pts** — You and your partner did not create a Kaggle page for your dataset.
Example analysis notebook scope and organization	**15 pts** — The example analysis notebook is well organized and demonstrates at least five interesting patterns in the dataset using fundamental transformation and visualization tools. The notebook is written and annotated for an audience of new data scientists who have a basic understanding of {dplyr} and {ggplot2}. The notebook provides legitimate educational utility and could be used in an introductory data science course to illustrate basic wrangling concepts.	**12 pts** — The example analysis notebook is well organized and demonstrates at least three interesting patterns in the dataset using fundamental transformation and visualization tools. The notebook is written and annotated for an audience of new data scientists who have a basic understanding of {dplyr} and {ggplot2}.	**9 pts** — The example analysis notebook is somewhat organized, but fails to coherently demonstrate three interesting patterns in the dataset. The transformation and visualization code does not demonstrate best practices, or is written for a more advanced audience without sufficient annotation or narrative guidance. The notebook may confuse new data scientists more than it would help them.	**0 pts** — You and your partner did not create an example analysis notebook.