

Paper	Exemplary	Good	Acceptable	Needs Work	Inadequate
Introduction and Conclusion	25	20	15	10	5
	Clearly and concisely describes the question and why it is of interest. Describes the data and why it is relevant for answering the question. Sets up a clear roadmap for the rest of the paper. Includes interesting further questions and ideas for future research.	Good introduction to questions and data, but a roadmap for the rest of the paper lacking. Good summary, but doesn't pull pieces together into a cohesive whole. Interesting ideas for future research	Introduction and roadmap unclear and missing important details. Summary patchy, but some attempt at synthesis and development of ideas for future work.	Rote description of data. No context provided for the question or data. Repeats findings with no synthesis. No proposals for future work.	Fails to introduce questions and data of interest. Fails to summarise findings or ask more questions.
Graphs	20	16	12	8	4
<i>Data Visualizations</i>	Atleast five data visualizations that are clearly explained. Captions are informative, axes are properly labeled, plotting symbols and colors are judiciously selected. All graphs serve a clear purpose. The graphs have variety and do not exhibit the same pattern. (Ex. All histograms based on features)	Five data visualizations, each enhance the paper. At Least two different types of graphs. A few have minor problems: e.g., inappropriate aspect ratios, poor labels, poor quality when printed.	Atleast three data visualizations that are explained. Some of the graphs are poorly presented. E.g. no labels, too many decimal places, poorly chosen aspect ratio etc.	Data visualizations are unclear and lacking. They do not support the question you framed and have fundamental flaws.	There are no good visualizations, or the visualizations are irrelevant to the framed question. Graphs do not support questions and findings.
Main Body	25	20	15	10	5
<i>Rigor</i>	Intense exploration and evidence of many trials and failures. You have looked at the data in many different ways before coming to your final answer. You have gone beyond what was asked: additional research from other sources used to help understand/explain findings. Your explanation is creative.	Plenty of exploration and investigation. Some additional research helps explain findings, and some of your ideas are creatively presented and explained.	Some exploration, but little evidence that you have selected the best of many ideas. Little or no additional research beyond your initial dataset.	You have done the bare minimum that was asked. There is no evidence to suggest that you tried multiple approaches (tables, graphics, or wrangling) before coming to your final conclusion.	There is no evidence of exploration. You merely present the data as-is with no clear narrative in mind.
<i>Questions</i>	All questions are clearly answered. Answers coherent and sufficient.	Questions are mostly answered. Some confusion or missing information.	Questions are addressed, but confusing and missing relevant information.	Questions are poorly answered. No sufficient content.	Questions are not answered.
Quality	5	4	3	2	1
<i>Text</i>	English is polished, concise, and clear. No grammar or spelling mistakes.	Clear and concise, but not elegant. A few spelling and grammatical errors.	Readable, but excessively verbose, or lacking in detail. A number of errors in text.	Marginally readable. Many errors.	Barely readable. Many spelling and grammatical errors. No evidence of proofreading.

Coding	Exemplary	Good	Acceptable	Needs Work	Inadequate
Data Cleaning	20	16	12	8	4
	Data cleaning techniques taught in the class are clearly utilized. All categorical variables are utilized through a relevant one-hot-encoding scheme. Null values are clearly addressed in a rational way, and numerical variables are standardized.	Most of the data cleaning techniques taught in the class are clearly utilized. Null values are addressed, one-hot encoding is used.	Some data cleaning techniques taught in the class are clearly utilized. One hot encoding or standardization of data attempted, but not done correctly	Data cleaning is very minimal. One technique taught in class is used inefficiently.	No data cleaning was attempted.
Coding Style	10		5		0
Naming Convention	Code follows a uniform naming convention. For example, all functions are in camelcase while other objects are in undercase with underscores separating words. Names are informative.		Naming convention is fairly consistent with some minor inconsistencies (1-3 occurrences). Names are generally informative.		No clear naming convention followed or is inconsistent in more than 3 places. Names are not carefully chosen.
Formatting & Comments	Functions and expressions are always organized concisely. Effective use of whitespace to make code readable. Code has clear and informative comments. Easy to read and understand. All features have a description.		Code has comments that partly help the reader to understand the code or partially describe the features. Difficult to read at times.		No comments. Code is redundant and hard to understand.
Organization	Project is clearly organized and follows the data science life cycle. If necessary, the execution is straightforward and requires that the reader only run a few lines of code to assess the code, and test the model.		Some organization of the code. May not fully follow data-science lifestyle. The user must modify the code to get the project to run.		No organization was attempted. Execution requires that the reader heavily modify the code..
Technical Mastery	15	12	9	6	3
	Code is compact and efficient (in terms of verbosity, not computation time). No unnecessary copy-paste when a suitable function could take care of the task. The code clearly demonstrates mastery of topics covered in this class rather than "hacks" used to patch problems.	Code demonstrates competence of major topics covered in this class. Presence of 1-2 "hacky" solutions that could have been solved in a better way.	Code shows a basic understanding of topics covered in this class. The code works, but is inelegant in more than 2 places.	Code reveals holes in basic understanding of topics covered in class. Code doesn't work in 1-3 places.	Code only demonstrates very basic functionalities. Lots of missing parts. Code doesn't work in at least 3 places.

Methodology	Exemplary	Good	Acceptable	Needs Work	Inadequate
Reasoning/Understanding of Core Methodological Concepts	20	16	12	8	4
	<p>Features are solid, innovative, and show research for the particular topic.</p> <p>All predictors are fully trained and evaluated.</p> <p>Core DS techniques such as cross-validation, and learning-test data split used correctly.</p> <p>Statement of the model and assumptions they are using for inference are all valid.</p>	<p>Good set of features. .</p> <p>All predictors fully trained.</p> <p>Core DS techniques such as cross-validation, and learning-test data split used correctly.</p> <p>Statement of the model and assumptions are clear.</p>	<p>Naive features..</p> <p>Most of the predictors trained.</p> <p>Attempted to use core DS techniques, such as cross-validation, and learning-test data split , but used incorrectly.</p> <p>Assumptions made for models are inaccurate or inferences made are inaccurate.</p>	<p>Very few attempts to use core DS techniques, such as cross-validation, and learning-test data split.</p> <p>Model assumptions are not clear and inferences are inaccurate.</p>	<p>No or very limited features.</p> <p>Incorrectly trained models.</p> <p>No attempts to use core DS techniques, such as cross-validation, and learning-test data split.</p>

Performance	Exemplary	Good	Acceptable	Inadequate
How “good” is your model	15	10	5	0
	<p>The approach to solving the question framed by the group provides <i>superior</i> performance than the baseline. (Eg. prediction task = predict majority, regression, predict)</p>	<p>The approach to solving the question framed by the group provides reasonably <i>better</i> performance than the baseline.</p>	<p>The approach to solving the question framed by the group provides many trials, but no improvement to performance than the baseline..</p>	<p>No or very <i>limited</i> performance gain. No evidence of significant effort.</p>