



Evaluation Rubric for Technical Report

Data Mining and Wrangling 1

Grading Procedure

The score assigned to a component is the lowest among the scores whose description is satisfied. For example, the descriptions for the Data Source component of the Completeness dimension are shown below.

Dimension	Component	Deficient (3 points)	Fair (6 points)	Good (8 points)	Excellent (10 points)
	Data source	Does not provide either high-level description or how they were collected by the source for any dataset	Describes how they were collected by the source but no high-level description for some datasets	Provides a high-level description of the dataset but not how they were collected by the source for some datasets	Provides a high-level description and how they were collected by the source for all datasets

Suppose there were two datasets. A high-level description was provided for dataset A, but how it was collected by the source was not provided. On the other hand, how dataset B was collected by the source was provided, but there was no high-level description of it. The score cannot be Excellent because a high-level description and how they were collected by the source were not provided for *all* datasets, in fact, none. It cannot be Deficient because either a high-level description or how they were collected by the source was provided for each dataset. Dataset A satisfied the score of **Good** whilst dataset B satisfies the score of **Fair**. The resulting score for this component would be **Fair** (6 pts) because the lowest score should be picked.



Evaluation Rubric for Technical Report

Data Mining and Wrangling 1

Writing Performance

Dimension	Component	Deficient (3 points)	Fair (6 points)	Good (8 points)	Excellent (10 points)
Completeness	Title	The title is too long, misleading, and inaccurate, or missing.	The title captures somewhat the study, but has inaccuracies.	The title captures what the study is about, but it can be improved.	The title succinctly reflects what the study is all about.
	Abstract	Several essential parts of the abstract are missing, or there is no abstract.	Provides an overview of the study, but fails to mention an important detail or multiple minor details.	Provides an overview of the study, but misses one minor detail.	Provides a complete overview of the study: context, problem statement, methodology, results, and recommendation.
	Problem Statement	The problem statement is not apparent to the reader and is not explicitly stated in the Introduction.	The problem statement is apparent to the reader but not explicitly stated in the Introduction.	The problem to solve is explicitly but not clearly stated in the introduction.	The problem to be solved is explicitly and clearly stated in the introduction.
	Motivation	It is not apparent why the problem should be solved, and there is no explicit motivation statement to justify it.	The motivation for solving the problem is apparent in the report, but is not explicitly stated in the introduction.	Provides an explicit but unconvincing motivation to solve the problem in the introduction.	Provides an explicit and convincing motivation to solve the problem in the introduction.
	Data source	Does not provide either high-level description or how they were collected by the source for any dataset.	Describes how they were collected by the source but no high-level description for some datasets.	Provides a high-level description of the dataset but not how they were collected by the source for some datasets.	Provides a high-level description and how they were collected by the source for all datasets.
	Data Exploration	Does not describe the raw data properly and does not provide simple, but relevant statistics for some raw data.	Does not describe the raw data properly but provides simple, but relevant statistics for some raw data.	Describes the raw data properly, but does not provide simple, but relevant statistics for some raw data.	Describes the raw data properly and provides simple but relevant statistics for all raw data.
	Methodology Overview	No implicit or explicit pipeline overview of the methodology is provided.	An implicit pipeline overview of the methodology is provided.	Provides an explicit but incomplete pipeline overview of the methodology.	Provides a complete and explicit pipeline overview of the methodology.

	Methodology Overview Details	There is no explicit methodology overview.	The methodology overview is specific, but not high-level.	The methodology overview is high-level, but not specific.	The methodology overview is high-level but specific.
	Methodology Details	Information about the methodology is not provided in the text.	Provided detail is insufficient to reproduce the pipeline at a high level.	Sufficient detail is provided to reproduce the pipeline at a high level, but not enough details to reproduce the results.	There is sufficient detail in the main text to allow a peer to reproduce the results.
	Results and Discussion (x2)	Does not present any insights at all.	Presents interesting insights, but some of them are not derived from the results.	Some interesting insights were not explicitly discussed, but all presented insights are backed by the results.	Discusses interesting insights that are all backed by the results (not necessarily technical).
	Conclusion (x2)	The results do not answer the problem.	The answer or solution to the stated problem can be completely derived from the results, but not explicitly stated.	The answer or solution to the stated problem is explicitly stated, but not clearly and accurately.	The answer or solution to the stated problem is explicitly, clearly, and accurately stated.
	Recommendations (x2)	Limitations and possible recommendations are noticeably absent from the report.	Limitations and possible future work were not explained and lack important details.	Limitations and possible future work were enumerated, but can still be further explained.	Provides a discussion of the limitations of the study and potential future work in detail.
	References	No or incomplete in-text citations and no properly formatted reference section	Has complete in-text citations, but the reference section is not properly formatted or incomplete.	Has properly formatted and complete reference section but incomplete in-text citations	All references and sources are properly cited in-text with properly formatted reference section following APA format.



Evaluation Rubric for Technical Report

Data Mining and Wrangling 1

Writing Performance

Dimension	Component	Deficient (3 points)	Fair (6 points)	Good (8 points)	Excellent (10 points)
Presentation	Organization	The document lacks a coherent flow and is hard to read. Most section headers are absent.	Some sections or paragraphs are disjoint. Important section headers are absent.	Most of the paragraphs are well-organized. Sections and headers are usually logical.	Paragraphs are organized. Sections and headers are logical and ease navigation.
	Aesthetics (x2)	Most visualizations were poorly generated and failed to integrate with the text.	Some visualizations were poorly generated and/or failed to integrate with the text.	Visualizations were appropriate and connect with the text but can still be improved.	Visualizations are creative, professional, and strongly integrates with the text.
	Figures, Tables, and Equations	Most elements lack labels and/or captions and may not be referred to in the main text.	Some elements lack labels and/or captions, but most are correctly captioned and referred.	All elements have labels, but some captions or descriptions can be made to be self-containing.	All elements have labels, self-contained captions, and are referred to in the text.
	Mechanics	Too many grammar and spelling errors. The report is poorly written and is very hard to read.	There are obvious grammar and spelling errors. The report can be improved by proofreading.	The report is mostly well-written. There are few grammar and spelling errors.	The report is well-written. There are no grammar or spelling errors. The text is easy to digest.



Evaluation Rubric for Technical Report

Data Mining and Wrangling 1

Technical Performance

Dimension	Component	Deficient (3 points)	Fair (6 points)	Good (8 points)	Excellent (10 points)
Data Analysis (x4)	Correctness of Methodology/Plan	The presented methodology does not address the stated problem(s) or there is no presented methodology.	The presented methodology correctly addresses most of the stated problem(s), but there are errors.	The presented methodology correctly addresses the stated problem(s) but not completely.	The presented methodology correctly and completely addresses the stated problem(s).
	Execution of Methodology/Plan	The presented methodology was not properly executed, or there is no methodology presented.	The presented methodology was mostly properly executed.	The presented methodology was properly executed but incomplete.	The presented methodology was completely and properly executed.
	Data Pre-processing	No detail or justification is provided for all filtering steps, and no number or proportion of the resulting data points is provided in all filtering steps as well. No investigation was done to determine why null values exist and how null values were handled were not detailed or sufficiently justified.	The number and proportion of resulting data points of each data filtering step is explicitly stated, but a step is not sufficiently detailed and justified. An investigation was done to determine why null values exist, but imputation or dropping of nulls is not detailed or not sufficiently justified.	Each data filtering step is sufficiently detailed and justified, but the number and proportion of resulting data points of a step are not explicitly stated. Imputation or dropping of nulls is detailed and sufficiently justified, but no investigation was performed to determine why null values exist.	Each data filtering step is sufficiently detailed and justified, and the number and proportion of resulting data points are explicitly stated, or no filtering was done. Imputation or removal of nulls is detailed, sufficiently justified, and performed only after investigating why null values exist, or there were no null values in the data.
	Depth of Analysis	The report presents a shallow analysis with no clear insights or description of the data.	The report presents sufficient analysis to partially answer the main research questions, but leaves some areas of further analysis unexplored.	The report explores the data to sufficient depth to provide an answer to the overarching research question. Includes minor extensions to original EDA based on initial results.	The report fully answers the overarching research question to theoretical and technical depth. Additional exploration warranted by initial results is carried out to completion.
	Accuracy and Mastery	Many errors in the facts and/or concepts are found in the report. The author has grievous misconceptions about the topic.	Some errors are apparent in the facts and/or concepts in the report. The author is hesitant to explain important topics.	No errors in the facts and/or concepts written in the report and the author clearly understands the topic.	There are no errors in the facts and/or concepts written in the report and the author shows mastery in a significant way.

	Significance and Impact	There is no clear value for the results and the analysis of the study.	The significance of the analysis and results is not apparent and/or does not have any clear application or actionable insights.	The author presents an analysis and results that drive and provide clear applications and actionable insights in the context of the study.	The author presents analysis and results that drive and provide clear applications and actionable insights beyond the context of the study.
--	-------------------------	--	---	--	---



Evaluation Rubric for Technical Report

Data Mining and Wrangling 1

Technical Performance

Dimension	Component	Deficient (3 points)	Fair (6 points)	Good (8 points)	Excellent (10 points)
Code Hygiene	PEP8 Compliance (x2)	Most of the code do not comply with the PEP8 style guide.	There are some apparent and obvious PEP8 violations. Most of the codes are still PEP8 compliant.	Minor PEP8 violations. Some code can be further improved to clearly adhere to the PEP8 style.	All code complies clearly with the PEP8 style guide.
	Code documentation	Most if not all functions have no docstring or most pieces of code lack important one-line comments or docstrings.	Some functions lack docstrings, or some pieces of code lack important one-line comments.	All functions and codes are completely documented with meaningful one-line comments or docstrings, as applicable, but some parts can still be annotated.	All functions and codes are completely documented with meaningful one-line comments or docstrings, as applicable.
	Nomenclature and Readability	Variables are named inconsistently and unintelligibly.	Most of the variables are named consistently. Some have unintelligible names.	Variables are named consistently, but their naming can still be improved.	Variables are named consistently and intelligibly.
Compliance to Requirements	Dataset, Format, Executability, and Anonymity	Three or more compliance requirements are not satisfied.	Two compliance requirements are not satisfied.	One compliance requirement is not met.	Uses at least one of the allowed datasets. All relevant files are provided except data files. Main notebook runs from top to bottom. Author cannot be identified with the submission.