STAT 10118/94: Pathways in Data Science, Summer 2021
Final Project Rubric

Group name/number:
Group members:
Project grade: /33

Criterion	Poor (0)	Needs Improvement (1)	Good (2)	Excellent (3)
Choice of research question and dataset.	The choice of question to investigate is ill-posed and unscientific, with no potential for interesting results.	The choice of question to investigate is sound, but may be trivial or uninteresting. The dataset is relevant to the question, but the report struggles to explain the content and source of the data.	The choice of question to investigate is sound, nontrivial, and interesting with potential for robust results. The dataset is well-suited to the question posed, and the report demonstrates a clear understanding of the content and source of the data.	The choice of question is particularly unique and interesting with potential for deep, unexpected insights. The dataset is well-suited to the question posed, and the report demonstrates a clear understanding of the content and source of the data.
Introduction and Background Exposition	The report does not pose or contextualize the question.	The report struggles to clearly pose and contextualize the question to be investigated. The reader is left confused about what the rest of the paper will be about.	The report poses and contextualizes the question to be investigated, but some things are left unclear or ill-defined, or there are few outside references for background information. The reader is left with a rough idea where the rest of the paper is headed.	The report clearly poses and contextualizes the question to be investigated and references several outside sources for background information. The reader is left excited to read the rest of the paper, with a clear conception of where it will be headed.
Data	Data is nonsensical.	Data is cleaned, appropriately normalized, or otherwise	Data is cleaned, appropriately normalized, or otherwise	Data is cleaned, appropriately normalized, or otherwise

		manipulated with many errors.	manipulated appropriately with a few errors.	manipulated appropriately with no errors.
Exploratory Analysis	Exploration of data is absent.	Exploration of data is present but weak or ill-explained.	Exploration of data provides insight and motivation for what to explore further, but with few issues or errors.	Exploration of data is excellent with strong insights and no errors and clearly motivates the following analysis.
Predictive Analysis (point values doubled)	Predictive analysis is absent or extremely hard to understand.	Predictive analysis uses a model, but the model is inappropriate or the justification for using the model is weak. Methods not covered in class are used "out of the box" with little to no explanation of the underlying mathematics. Error metrics are not reported. Many mistakes in reasoning are present.	Predictive analysis of data has a clearly stated model and training procedure, as well as a basic understanding of the underlying mathematics if a method not covered in class is used. The method used is appropriate for the questions posed and a justification is provided. Error metrics are appropriately reported. Few mistakes in reasoning are present.	Predictive analysis of data has a clearly stated model and training procedure, as well as a superb understanding of the underlying mathematics if a method not covered in class is used. The method used is appropriate for the questions posed and a justification is provided. Error metrics are appropriately reported. There are no mistakes in reasoning.
Visualizations	VIsualizations are absent or not explained at all.	Visualizations are present, but poorly labelled or unexplained.	Visualizations are present, with axes and figures appropriately labeled. Visualizations contribute to the understanding of the data analysis, though there might be some errors.	Visualizations are free from technical issues and offer deep insight into the research question. Particular effort has been made to make plots look "pretty".
Conclusion	Conclusion is absent or not relevant to the	Some answers to the hypothesis are given and	Answers to the hypothesis are given and	Answers to the hypothesis are given and

	predictive analysis.	weakly supported by the results of the predictive analysis. There is limited further discussion, or many mistakes in reasoning.	supported by the results of the predictive analysis. There is some discussion about the successes, limitations, and failures of the results. There may be a few mistakes in reasoning.	well-justified by the results of the predictive analysis. There is a convincing and nuanced discussion about the successes, limitations, and failures of the results. There are no mistakes in reasoning.
Jupyter Notebook	Jupyter Notebook is unreadable or extremely disorganized	Jupyter Notebook is messy and hard to follow. It is difficult to see which parts of the Notebook correspond to which parts of the report.	Jupyter Notebook is mostly organized and easy to read. It is mostly easy to see which parts of the Notebook correspond to which parts of the report, with possibly a few confusing cells.	Jupyter Notebook is extremely organized and readable and contains ample exposition relating it to the report.
Presentation	Presentation is extremely difficult to follow and understand.	Students struggle through presentation and seem to have a weak understanding of the content.	Presentation is mostly clear and easy to follow, demonstrating a fair understanding of the material.	Presentation is very clear and easy to follow, demonstrating a deep and confident understanding of the material. Audience questions are well-answered.
Report formatting	Report is very poorly or unprofessionally formatted, or very difficult to read due to a large number of technical errors.	Report is somewhat sloppy, but not to the point where it significantly affects the reader's ability to understand.	Report is neat and cleanly formatted, possibly with a few technical errors.	Report is exceptionally well-formatted and a pleasure to read.