

# Stat 172 Final Project (Individual)

Dr. Lendie Follett

2024-11-04

## Data Science Project Management

The goal of this project is to:

1. encourage you to build a data science portfolio
2. allow<sup>1</sup> you to do it in a way that interests you
3. provide you with experience communicating your work with other data scientists through a README file.

At a minimum, you are being asked to write a README file to aid other data scientists in reproducing and understanding your work. You can receive full points for doing this in regards to your Wesley Life group project (though the README should be written individually). Some of you may wish to instead focus on a different project that you want to showcase in your online portfolio. That is welcomed and encouraged. In any case, I encourage, but do not require, you to make this a part of a publicly available Github repository.

## RUBRIC

Your README file should be either a plain text (.txt) file or a markdown (.md)<sup>2</sup>. It should have the sections/features/qualities outlined in the rubric below.

There are many examples of README files on your favorite Github repositories. Here's one to give you an idea of (1) what a .md file is and (2) what you might want to put in it: <https://github.com/LendieFollett/Hybrid-Targeting>.

## Project Overview

- **Outstanding (10):** Clearly describes project goals, methods, and data sources in a succinct summary. Conveys purpose and utility of the project, making it easy for others to understand at a high level.
- **Proficient (5):** Provides an overview, but lacks depth or specificity about methods or data sources. Gives a general idea of the project's purpose but may require clarification.
- **Needs Improvement (0):** Missing or vague project overview; does not clearly convey purpose or context. May leave the reader confused or unclear on the goals of the project.

## Data Sources and Preparation

- **Outstanding (10):** Provides comprehensive details on data sources, format, and preparation steps. If data is external, offers clear access instructions. Data processing is done via code. Any non-code pre-processing is detailed. Lists all needed packages, with instructions on setting up the environment. Instructions are complete and error-free, enabling seamless setup.
- **Proficient (5):** Mentions data sources and formats, though lacks details on access, pre-processing, preparation, or packages. Users may need to infer or explore further to fully understand data handling.

---

<sup>1</sup>but not require

<sup>2</sup>note that this is different than a RMarkdown file - a .md file will automatically be rendered in Github while a .Rmd will not.

- **Needs Improvement (0):** Data sources, format, and preparation instructions are missing, incomplete, or incorrect. Reader is left unsure about data handling or preparation steps.

## Code Execution

- **Outstanding (10):** Gives clear, step-by-step instructions on how to execute the code. Lists required commands and settings, with examples. Code is ready for reproduction by others.
- **Proficient (5):** Code execution instructions are present, but minor steps or settings may be missing. Most users will be able to execute with some effort to clarify steps or assumptions.
- **Needs Improvement (0):** Instructions are vague, incomplete, or missing. Readers would struggle to run the code or reproduce results without substantial additional guidance.

## Explanation of Methods

- **Outstanding (10):** Provides concise, accurate descriptions of key analytical methods and algorithms used. Discusses the reasoning behind method selection in the project's context. concise: (3-5 sentences paragraph)
- **Proficient (5):** Mentions the main methods used but lacks context or clarity on why specific approaches were chosen. Methods are identifiable but not fully explained.
- **Needs Improvement (0):** No description or very limited explanation of methods used. Lacks transparency on key steps or reasoning, leaving readers unclear on the analytical approach.

## Professionalism and Clarity

- **Outstanding (10):** README is well-organized, free from grammar and spelling errors, and professional in tone. Uses technical language appropriately for a data science audience.
- **Proficient (5):** README is generally clear but may have minor grammar or spelling errors. Tone and language are professional, though some technical details may lack clarity or precision.
- **Needs Improvement (0):** README is difficult to follow, contains grammar/spelling errors, or is unprofessional in tone. Technical language is unclear or overly simplistic.

FYI, the following is often included in README files, though I won't grade you on this part:

## File and Folder Structure

- List all files and folders (as they are in your project, which may or may not be on Github) with descriptions of their purpose and relationship to one another. Instructions make it easy to navigate and understand the organization.