

Case Study Rubric — DS 4002

Due: TBD

Submission Format: Upload link to GitHub repo to Canvas and submit a hard copy PDF in-person

Type of Assignment: Individual Assignment

Preparatory Assignments: Everything in the course

Why am I doing this? This case study is an opportunity for you to engage in an individual assignment where you can fully utilize your conceptual and technical skills. This project will be more realistic, tailored to a real-world scenario rather than a typical, clear-cut class assignment.

What am I going to do? Throughout this course and your academic career, you have honed your data science skills through classes and group projects. This case study is a chance for you to display these skills in a specialized project. The deliverables that you will submit include:

- GitHub repo link, including all of your code, data, PDFs, extra materials, and sources
- Hard copy of the written portion

How will I know I have succeeded? You will meet expectations on this case study when you have followed the criteria in the rubric below:

| Formatting | <ul style="list-style-type: none">• Topic — The topic for this case study is analyzing national COVID-19 data from 2020 to early 2024• Repository — A GitHub repo that should contain the following items (use PDF for written documents):<ul style="list-style-type: none">○ The hook document used to entice the data scientist○ The rubric created for reproducing the case study○ The master script(s) containing all of the code needed to produce the final deliverables<ul style="list-style-type: none">■ Be sure to thoroughly document the code■ There are two datasets that should be cleaned, merged, and subsetting■ The model should be trained and tested on the cleaned data■ The model should then be used to |
|------------|--|

| | |
|---------------|--|
| | <p>time-series forecasting prediction on future counts of COVID-19 in the United States</p> <ul style="list-style-type: none"> ○ The original data from the CDC ○ The sources for background information (IEEE formatting) <ul style="list-style-type: none"> ● Written Portion — A high level, overarching concept of the case study. It should contain the following items: <ul style="list-style-type: none"> ○ The hook document (one page) ○ Two printed sources of background information on the topic |
| Hook Document | <ul style="list-style-type: none"> ● Set up the data scientist with some background on the topic as well as the basic outline for the project ● <u>First Paragraph</u>: Motivation behind the nature of the project and additional information for context ● <u>Second Paragraph</u>: Overall explanation of the deliverables, saving the details and specifications for the rubric |
| Rubric | <ul style="list-style-type: none"> ● Explain, in detail, the motive of the project, the tasks that follow, and what is needed to successfully reproduce the results |
| Materials | <ul style="list-style-type: none"> ● GitHub repo ● All code necessary to reproduce the results ● The original data (will be cleaned within the code) ● All sources, including datasets and background information |