

AQI Case Study Rubric

DS 4002 - Spring 2024 - Instructors: Javier Rasero and Harsh Anand

Due: see canvas

Submission format: GitHub repository (submitted by link to canvas)

Group Assignment

General Description: Submit to Canvas a link to your group's GitHub repository for this project. This should include data sources, code, and more. You are also responsible for creating and delivering a presentation.

Why am I doing this? The goal of this assignment is to develop and demonstrate data science technical skills, presentation skills, and communication skills. This project will give students a deeper understanding of predictive modeling, time series data, and written/oral communication. The goal of this project is to inspire students with the real world, ever important, application of Data Science. This project will help set students apart in the professional world.

- Course Learning Objective: translate processes from theory in the scientific method into functioning data science pipelines

What am I going to do? You will work together as a team to build a thorough analysis plan, identify necessary supplemental data sources, develop and implement code + data science algorithms, conduct additional analysis, and compile this together in a final deliverable presentation. Teammates may each choose to work on a select part, or teammates may collaborate across all elements of the project. The end deliverable should be cohesive and representative of everyone's ideas/work.

Tips for success:

- Focus on developing a thorough analysis plan and understanding of the prompt and relevant data sources.
- Work together, bouncing ideas off of one another to develop a more comprehensive analysis of the time series data.
- Try different approaches/different methods, creativity and innovation is important.
- Document code well and create an appropriate, easy to follow file management system.
- Development of the presentation should be tailored to non technical as well as technical audiences, focusing on your audience.
- Connect the project to real life, using examples and tying back to policy/societal implications of your findings.

How will I know I have succeeded? You will meet expectations on AQI Case Study when you follow the criteria in the rubric below.

Requirements	<ul style="list-style-type: none"> • One Github Repository (submitted via link on canvas) • One Presentation (May be Word, Powerpoint, PDF, etc.)
Github Repo	<ul style="list-style-type: none"> • This should include: <ul style="list-style-type: none"> ○ A README.md file ○ A LICENSE.md file ○ A SCRIPTS folder ○ A DATA folder ○ AN OUTPUT folder
README.md	<ul style="list-style-type: none"> • <u>Goal</u>: This file serves as an orientation to everyone who comes to your repository, it should enable them to get their bearings. • Section 1: Software and platform section • Section 2: A Map of your documentation. • Section 3: Instructions for reproducing your results.
LICENSE.md	<ul style="list-style-type: none"> • <u>Goal</u>: This file explains to a visitor the terms under which they may use and cite your repository. • Select an appropriate license from the GitHub options list on repository creation.
SCRIPTS folder	<ul style="list-style-type: none"> • <u>Goal</u>: This folder contains all the source code for your project. • Include all the scripts you used, naming in order of execution, and including lots of comments for reproducibility.
DATA folder	<ul style="list-style-type: none"> • <u>Goal</u>: This folder contains all of the data for this project. • Include the initial data, and the final data analyzed. • Contain code in scripts file if necessary to convert initial data into final data. • If your data does not fit in GitHub use a single file explaining the process to obtain the dataset. • A Data Appendix file as a PDF, which will include text that you type, as well as tables, figures, and other descriptive statistics.
OUTPUT folder	<ul style="list-style-type: none"> • <u>Goal</u>: This folder contains all of the output generated by your project, e.g. figures, tables, etc. • Tables, figures shown in your presentation should be here.
Presentation	<ul style="list-style-type: none"> • About 10 slides (include slide numbers) <ul style="list-style-type: none"> ○ Title & Outline ○ Motivation/Context/Hypothesis/Research Question/Modeling Approach/Goal/Etc. ○ Data Explanation/Acquisition ○ Analysis Plan and Justification ○ Tricky Analysis Decision ○ Bias and Uncertainty Validation ○ Results/Conclusions

	<ul style="list-style-type: none"> ○ Next Steps ○ References/Resources/Acknowledgements ○ Closing Slide
References	<ul style="list-style-type: none"> ● All references should be listed at the end of the document ● Use IEEE Documentation style (link)

Acknowledgements: Special thanks to Jess Taggart from UVA CTE for coaching on making this rubric. This structure is pulled from [Streifer & Palmer \(2020\)](#).