

DS 4002 – Spring 2023 - Victoria Feist

Submission format: Upload .pdf slides and github repository link

Individual Assignment

General Description: Submit a link to your repository and .pdf of your slides.

Preparatory Assignments – Case study description and attached resources.

Why am I doing this? As data scientists, we often must address complex questions and communicate our findings in a digestible manner to those outside of our field. Since problem-solving and science communication are vital to data science careers, it's important to practice executing a project from start to finish. Placing yourself in the shoes of a career data scientist addressing a problem you feel passionate about will allow you to model what being a data scientist is really like, and allows you to be prepared once you actually enter the workforce.

- Learning Objective: use critical thinking to devise a method of analysis
- Learning Objective: seamless communication of coding (via annotated markdown and organized github repository)
- Learning Objective: present findings in an understandable manner

What am I going to do? In order to begin, read the file named “CS2_Hook.” This will explain the prompt of the assignment in an exciting manner in order to encourage you to place yourself in the shoes of a data scientist. This document will also introduce what you will need to do to complete this case study. Next, read the provided resources on how to begin with the code. You will then obtain and clean the data, perform your analysis, and prepare a presentation to communicate your conclusions. All code and resources used should be uploaded to a singular Github repository.

Deliverable

- Github repository - to provide all annotated code, figures, and resources so that your results may be replicated
- Presentation - to ultimately communicate your findings to a group of non-data scientists

Tips for success:

- Be adventurous and persevere. You may not have ever coded a model like this before. Don't be scared! Encourage yourself to keep trying if you don't immediately succeed.
- Keep it simple. Your final audience will be a group of policymakers, not data scientists. They will only need to know the important information related to the task at hand.
- Stay organized. The dataset at hand is large, and the method of analysis may be complex.

How will I know I have Succeeded? You will meet expectations on this case study when you meet the requirements listed below.

Spec Category	Spec Details
Formatting	<ul style="list-style-type: none"> ➤ Github Repository <ul style="list-style-type: none"> ○ Should contain a source code folder, data folder, figures folder, license declaration, readme.md, and a copy of your slides. ➤ Presentation <ul style="list-style-type: none"> ○ Should contain 8-12 slides as a .pdf
Github Repository	<p>Goal: this repository is a credible method of explaining your results. Someone should be able to download this repository and produce the same finding and understand how to do so well. (You're working for a government organization!)</p> <ul style="list-style-type: none"> ➤ Source code folder <ul style="list-style-type: none"> ○ Should contain an annotated markdown file, including packages used. Do NOT include unused packages in your final file. ➤ Data folder <ul style="list-style-type: none"> ○ Should contain your dataset (pre- and post-cleaning) ➤ Figures folder <ul style="list-style-type: none"> ○ Contains a downloadable version of all of your figures with captions ➤ License declaration <ul style="list-style-type: none"> ○ In the form of a license.md file (use MIT license) ➤ Readme.md <ul style="list-style-type: none"> ○ Should explain how to use the repository. Sections should be designated by ## ○ Includes: <ul style="list-style-type: none"> ■ A table of contents ■ A brief statement of purpose ■ A description of the source code folder ■ A description of the data (including a data dictionary) ■ A figures section including all figures and captions ■ A references section ➤ References <ul style="list-style-type: none"> ○ Attach any new references you used to complete this task. ➤ Copy of slides

	<ul style="list-style-type: none"> ○ In .pdf format
Presentation	<p>Goal: this presentation will be your primary method of communicating your findings to other government officials. It will allow you to alert the officials of the crisis at hand in a succinct manner.</p> <p>This presentation should include the following slides:</p> <ul style="list-style-type: none"> ➤ Title (1 slide) <ul style="list-style-type: none"> ○ List your primary finding as the title; include your name and the date of this presentation ➤ Motivation (1-2 slides) <ul style="list-style-type: none"> ○ Using the provided article resources, explain the motivation/context for your research. Alert the group of the inefficacy of current policies using cited sources. ➤ Hypothesis (1 slide) <ul style="list-style-type: none"> ○ State your hypothesis ➤ Data (2 slides) <ul style="list-style-type: none"> ○ Introduce your dataset. ○ Provide a data dictionary. ○ Explain how you cleaned the dataset. ➤ Analysis Plan (1-2 slides) <ul style="list-style-type: none"> ○ Explain to viewers how you are reaching this conclusion. However, avoid too much jargon. Explain in layman's terms, but include necessary details. ○ Example: instead of “I used an ARIMA model with the auto.arima package”, say “I created a model that can predict future trends in greenhouse gas emissions using data from previous years.” ➤ Analysis/Bias Justification (1 slide) <ul style="list-style-type: none"> ○ They may be skeptical of your findings! Remove any uncertainty your audience may have by thoroughly explaining any analysis decisions and potential biases ➤ Results (maximum of 3 slides) <ul style="list-style-type: none"> ○ Show your audience your predictive results! ○ Include any necessary figures and interpret your graphs. ➤ Conclusions (1 slide)

- | | |
|--|---|
| | <ul style="list-style-type: none">○ This is your take away slide! List any important findings using clear sentence structure and language.○ What do you want your audience to remember?➤ References and Acknowledgements (1 slide)<ul style="list-style-type: none">○ Cite your sources and any contributors. |
|--|---|