Case Study Rubric

DS 4002 – Spring 2024 - Nathan Geng

Individual Assignment

General Description: Submit a link to your GitHub repository and a PDF of your final presentation.

Preparatory Assignments – Read through the hook document and attached resources.

Why am I doing this? The purpose of this project is to present a complex question and your proposed solution in a practical and easily digestible manner to a board of healthcare professionals. In data science, there are many instances where your audience might not be familiar with the information or methods you use to solve the problem; it is your job to help them understand by communicating that information well.

- Learning Objective: prepare findings for presentation to a board of healthcare professionals.

What am I going to do? In the folder, read the document titled "CS3: Hook Document", which explains the case study scenario, motivations, and ultimate goals of the project. It will also help you in guiding where to start. Next, you can choose to read through some of the attached resources to become familiar with the concepts/methods/packages that will be used to complete the data analysis portion of the project. You will then obtain the dataset, perform analysis, and prepare a final presentation to present your conclusions. All of your code and analysis should be uploaded into a Github repository. Deliverables include:

- Github repository and data set – to provide resources like code and data so that results can be replicated
- Presentation - to communicate findings to a group of non-data scientists
- List of references - a compilation of websites/articles that you used in researching for the project

How will I know I have Succeeded? You will meet expectations on CS3 Create Case Study when you follow the criteria in the rubric below.

Formatting	o Repository o Presentation o References

Repository	● <u>Goal</u>: this repository is a means of both code storage and results reproduction. Someone should be able to download this repository, understand how to proceed with the project, and produce the same findings you did.
	    o README.md
	        ▪ Explains how to use the repository

	▪ Should include:
	   ● A description of the platform and add-on packages that you used to complete the project
	   ● A map of documentation that provides the outline for the rest of the repository, organized by folder name and their respective subfolders
	   ● A detailed set of instructions for results reproduction; someone should be able to read this and successfully attain the same results
	
	o LICENSE.md
	   ▪ Explains how a visitor may use and cite the contents of your repository (use MIT license)
	
	o SCRIPTS folder
	   ▪ Contains all of the source code for your project
	   ▪ Code should be annotated for the ease of potential viewers
	
	o DATA folder
	   ▪ Contains the pre- and post-cleaning dataset
	   ▪ Include a data appendix as a PDF, which contains all significant variables in the dataset and a short description
	
	o OUTPUT folder
	   ▪ Contains all of the outputs generated, like figures, tables, p-values
	   ▪ Use informative titles

Presentation	● <u>Goal</u>: this presentation should be the primary method of communicating your findings, in a succinct manner
	  o Should include:
	
	   ▪ Title (1 slide): include title, name, and date
	
	   ▪ Motivation for Project (1): explain the context of the project and potential areas of improvement
	
	   ▪ Hypothesis (1)
	
	   ▪ Data (1): Introduce the dataset and data appendix; explain data cleaning process
	
	   ▪ Analysis Plan (2): Explain the process with which you decided to undertake the project

	▪ Potential Bias (1): Explain any tricky analysis decisions and potential biases to address audience uncertainty ▪ Results/Conclusions (2): Present final findings ▪ Next Steps (1): Present on how you would like to proceed with the project with the given results ▪ Acknowledgements/References (1)
References	● Goal: Provide credit for the resources you used over the course of this project