Case Study – Influenza Vaccination Rate Modelling

DS 4002 – Fall 2024 - Instructor: Loreto Alonzi
Due: TBD

General Description: Submit a link to a repository including all inputs and outputs necessary for completion of project and creation of deliverables.

Preparatory Assignments: All previous Data Science curriculum and practice

Why am I doing this? Case studies are a great way to assess different skills and techniques gathered throughout the Data Science curriculum. By combining knowledge of datasets, analysis, modelling, and gathering key insights, students can generate concrete deliverables and takeaways from all kinds of data. In completing a case study, students get to practice the analysis process in a real world scenario, similar to one they might encounter later on in the workforce.

- **Course Learning Objective:** Analyze time scale data from
- **Course Learning Objective:** Determine key insights from analysis findings
- **Course Learning Objective:** Conduct relevant EDA to better understand data structure

What am I going to do? For this assignment, you will be reading through the contents of this folder to gain context of the problem and the necessary information required to develop a sound analysis. Be sure to thoroughly read through this rubric and the hook document to provide a full context for the type of data to analyze and the overall objectives of the analysis. You will have the opportunity to analyze data from the past, as well as creating a model to predict how the data of the past will impact future numbers.

Tips for success:

- **Be thorough.** Make sure to capture all the information available before beginning your analysis.
- **Remember the end objective.** Try not to get caught up in the details of the model, remember to keep the bigger picture in mind while simplifying each step if necessary.
- **Talk to the instructors.** This assignment includes modelling techniques that you may not have practiced in the past, reach out to your instructors if anything is unclear.
- **Have fun!** This project should not cause excessive levels of stress. If anything is unfamiliar to you, remember you have a network of data science professors and students where you can seek out support.

How will I know I have Succeeded? You will meet expectations on Case Study – Influenza Vaccination Modelling when you follow the criteria in the rubric below.

Spec Category	Spec Details

Formatting	• Repository: An online folder on GitHub.com containing the materials necessary for the conclusion of the project.
	  ○ Scripts:
	  ▪ Provide all of the code used to complete EDA and modeling analysis in an individual folder.
	  ○ Data:
	  ▪ Submit all data necessary for inputs into a folder in your GitHub repository.
	  ○ Outputs:
	  ▪ Compile all the outputs derived from your code into one document where graphs and charts can be easily accessed.
	  ○ README file
	  ▪ Include a README file to outline the contents of your repository.

Scripts	● Goal: This folder contains all the source code for your project. ● Include all the scripts you used. Try to name each script according to the order it needs to be executed to reproduce the results. ● All script files should include header comments at the beginning of a script to provide information that anyone working with or executing the script should be aware of. Throughout all your scripts, you should include copious comments explaining what each command or sequence of commands accomplishes and what the purpose is

Data	● Goal: This folder contains all of the data for this project.
	● You should AT LEAST the data include the initial data, and the final data analyzed.
	  ○ If needed, the code in the SCRIPTS folder should be able to get you from the initial piece of data to the final one. N.B. If the initial and final data are the same, then just include that dataset.
	● Place data directly in this folder if the file is small enough for GitHub
	● If your data is too large to fit in GitHub, write out detailed instructions outlining the process to download the dataset, including a link to the resource used to find the data.
	● A Data Appendix file as a PDF, which will include text that you type, as well as tables, figures, and other descriptive statistics.
	  ○ This file should be organized in sections, with a section for each dataset analyzed.
	  ○ Each section should begin with a statement of what the unit of observation is--that is, it should explain what kind of object each row of the data file represents.

	○ After that, you should include a subsection for each variable in the analyzed dataset.
Outputs	● Goal: This folder contains all of the output generated by your project, e.g. figures, tables, etc. ● The content here can be in progress when MI3 is complete. It should be finished during MI4 though. ● Importantly, any information like tables, figures shown in your presentation should be here. ● Use informative names for your files.
README File	● Goal: This file serves as an orientation to everyone who comes to your repository, it should enable them to get their bearings. ○ Use markdown headers to divide content. ○ Make an H2 (##) section explaining the contents of the repository ● Section 1: Software and platform section ○ The type(s) of software you used for the project. ○ The names of any add-on packages that need to be installed with the software. ○ The platform (e.g., Windows, Mac, or Linux) you used. ● Section 2: A Map of your documentation. ○ In this section, you should provide an outline or tree illustrating the hierarchy of folders and subfolders contained in your Project Folder, and listing the files stored in each folder or subfolder. ● Section 3: Instructions for reproducing your results. ○ In this section, you should give explicit step-by-step instructions to reproduce the Results of your study. These instructions should be written in straightforward plain English, but they must be concise, but detailed and precise enough, to make it possible for an interested user to reproduce your results without much difficulty.
References	● Include all references used that were not already given in the reference material in an individual reference file.

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