Case Study Rubric

DS 4002 – Spring 2024
Due: TBD

Individual Assignment

General Description: Submit to Canvas a link to your case study repository.

Why am I doing this? This is your opportunity to showcase your technical and conceptual skills while analyzing a topic of significant media coverage. As you work through this assignment, you will be exposed to the ways that data analysis can be used in a real-world context to gather information about public opinion and how people express their opinions on social media. Your conclusions from this project may help inform statistical patterns that occur when individuals are heavily scrutinized by the public on social media.

What am I going to do? You will utilize the csv files provided in the case study repository to perform a sentiment analysis, comparing the sentiment of Simone Biles and Jonathan Owens' Instagram comments. The case study has provided cleaned data (clean_sb.csv and clean_jo.csv) which you may use, or you may clean the provided jo.csv and sb.csv files yourself and use these for your analysis. You will use the VADER package in Python to carry out this analysis, and compile it into a github repository that includes a written summary of your research results and any takeaways you have gathered from your research.
Deliverables include:

- Github repository containing:
    - README.md file that provides a brief introduction of the project and includes a list of references (references both provided by the case study AND acquired by the student)
    - Source Code File including all Python code for the analysis
    - Utilized csv data files
    - One-page written summary of research results and takeaways

Tips for success:

- Thoroughly research the VADER package and familiarize yourself with Python. A link to the VADER package documentation has been provided on the case study repository, but your project will be most successful if you look at additional resources.
- Make your variable names easily interpretable and aligned with the purposes they serve to avoid confusion in your script.

How will I know I have Succeeded? You will meet expectations when you follow the criteria in the rubric provided:

Formatting	● Repository – A GitHub repo (and cloud storage folder if necessary) containing all materials. o Submit a link to the repo. o Everything is contained in the repo or linked to it if appropriate. o Contents ▪ README.md ▪ Source Code File ▪ Data Files ▪ Written Portion o Use pdf format when possible. o For code and data products use the appropriate format for whatever it is.
README.md	● Make an H2 (##) section explaining the contents of the repository . ● Provide a brief summary of what is included in your final deliverable, this does not have to be detailed but should provide people with enough information to become oriented to your repository. ● Include a list of references in IEEE citation style at the end of the file. o References both provided by the case study AND acquired by the student.
Source Code File	● This Python file should contain all of the code included in your analysis. ● Include an exploratory data to gauge distribution of Instagram comments across time. ● Then, complete your VADER sentiment analysis for each of the two datasets that you are using and compare the scores between the datasets. ● Your script should include header comments at the beginning to provide details that anybody working with the script should be aware of. ● Include copious comments explaining what each command or sequence of commands accomplishes and what the purpose is.
Data Files	● Include in the repo any csv files that you used (theoretically you would have used two of the provided csv files - either you used clean_sb.csv and clean_jo.csv OR sb.csv and jo.csv). ○ The README.md file in the case study original repository has more detailed information about what specifically is in the contents of the csv files. The files include the Instagram comment data and additional fields that you will be analyzing. ● It is recommended that you put both of your csv files into a 'DATA' folder for your own organizational purposes, but this is not required and points will not be deducted if you do not make a folder.

Written Portion	● One-page pdf formatted document. ● Explain how you carried out your exploratory data analysis and your sentiment analysis. ● Discuss your interpretation of the results. ● Reflect on any steps you could have gone about differently, what you've learned, and potential future directions in which to take your project. Are there any new questions that came about throughout this procedure?