CS-FOOTBALL TWEET Rubric

DS 1000 – Fall 2023 – Adina Kugler
Due: December 2023
Submission format: Link to github repository (collab assignments)

Individual Assignment

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

Preparatory Assignments – Class sessions about case study reading. Class sessions about using APIs in Python.

Why am I doing this? As a data scientist, it is important to learn how to answer research questions. You have the tools to explore your interests, and practicing with case studies is a great way to utilize your creativity and problem-solving skills essential to data science. In this case it is a fun example of a common argument with no real solution. This case study assignment is an individual assignment, but you are encouraged to ask questions, consult with your peers, and research. Research is an essential skill for this case study, and comparing the wide variety of results will allow for a comprehensive understanding of different approaches to a single problem. All in all, this project with make you a better data scientist.
- Course Learning Objective: logic and problem solving
- Course Learning Objective: creative thinking
- Course Learning Objective: communication of results

What am I going to do? To being, you will read a one-page introduction to the case study where you will be directed to the given task. While reading, think about how you would address the problem on a conceptual basis instead of a technical. Take time to jot down ideas. At the bottom of this introduction is details regarding the deliverables. Think about how your ideas could fit into these deliverables and the steps to get you there. The main deliverable is a github repo. Within the github repo, there should be three main components. There should be two one-page written documents discussing reasons for a particular part of the project. There also should be an example data set. Included in the repo should be any code and references you used for the deliverables. There are simple solutions to each problem, but no one is correct or incorrect. The main goal of this assignment is to be able to explain the choice you made. You will be trying to answer "how can you prove your point AND why?"

Tips for success:
- Be creative. The prompt is meant to allow for freedom of choices. Work with the fictional side of this. The statements do not need to be technical; instead make them fun.
- Do not fixate on being right. This assignment is specifically designed not to have wrong answers. You should be able to justify your work. The vagueness of the prompt is intentional so that you can define some of the criteria yourself.
- Be confident in your choices. There are millions of reasons you can choose that something works or doesn't, but make sure you stand with your decisions.

Helpful Hints about the Prompt:
- Remember you need data for all teams to prove you are better than them all
- The NFL or a team posting is not a fan
- Retweets may overcrowd rate limits and confound your results, same for copied tweets
- Time of pulling the tweets could have an indicator based on performance and not fans
- You may want to complete this project at any time of year
- You do not need to perform sentiment analysis, but the data needs to be set up for it
- Think of unique ways a team is identified
- You do not need to use outside resources, but they could be helpful
- You must figure out what fairness means to you and justify it
- You can choose a new NFL team, but provide a reason why (it can be personal)

How will I know I have Succeeded? You will meet expectations on CS-FOOTBALL TWEET Case Study when you follow the criteria in the rubric below.

Spec Category	Spec Details

Formatting	• Repository – A new github repository
	  • Create a new github repo for this assignment containing
	    • README.md
	    • Text Document 1
	    • Text Document 2
	    • CSV file
	    • LICENSE
	    • Images folder

README.md	• Goal: This file is what is an overview of the files that will be assessed for this assignment • Structure this file in a way similar to an executive summary and table of contents • Outline the purpose of the case study, context, and motivation • Include links to each portion of the case study including: Text Document 1, Text Document 2, sample Data, and references • Include a data dictionary • Include a short summary of your reasoning for your filter and time criteria with a fairness assessment (1-3 sentences) • Outline in one paragraph how your results could be considered fair • Include a paragraph regarding any limitations or challenges • Include any necessary information that was not present in the resources provided • If you did not follow the prompt, reason why you chose a different team
Text Document 1	• Goal: This is file 1/3 of what will be assessed for the assignment • Structure this file in such a way to be easily readable by an individual who has read the prompt • Include the filter criteria you chose to gather your data via Twitter at the top of the document

	• Include reasoning for why you chose to filter your data in this way • Include why you believe it was fair • Justify your choice in a one-page document • Include relevant reasons and sources • Include references • Text file format • No code is needed, it is about justifying your reasons
Text Document 2	• Goal: This is file 2/3 of what will be assessed for the assignment • Structure this file in such a way to be easily readable by an individual who has read the prompt • Include the time period you chose to gather your data via Twitter at the top of the document • Include reasoning for why you chose a specific time period for gathering your data • Include why you believe it was fair • Justify your choice in a one-page document • Include relevant reasons and sources • Include references • Text file format • If the time period is not available due to course timing restraints, explain when you would have pulled data and make it clear you did not pull data at the time you wanted • No code is needed, it is about justifying your reasons
Sample Dataset	• Goal: This is file 3/3 of what will be assessed for this assignment. It is as a dataset pulled from Twitter to prove the technical component of this task • Use a Twitter to pull data using your filter criteria • Pull data from Twitter during the time period your selected • If the time period is not available due to course timing restraints, pull any time • The dataset should be in CSV format • The row should be unique tweets • Columns should include the NFL team and the Tweet • Other columns are allowed but not required • There should be at least 15 unique tweets • Some data cleaning may be necessary
LICENSE	• Goal: Explain to readers the terms under which they may use and share your work. • The MIT license is the default recommendation
Additional Documents	• Goal: Include any or all additional documents needed to make your point • This could include information about sources, data cleaning processes, and code