Case Study Rubric

DS 4002 - Spring 2025
Submission Format: A link to a Github Repository.
Individual Assignment

Why am I doing this? This is your opportunity to showcase yout technical and conceptual skills to create a basic model in an area of relevant technology. As you work through this assignment, you will learn more about time text analysis, especially sentiment analysis and how this kind of analysis can be combined with other kinds of data analysis like time series analysis. You will be exposed to the ways in which these models can be utilized and evaluated to draw conclusions. You will then evaluate how useful these conclusions are.

What am I going to do? You will build a sentiment analysis model using te VADER package in python to analyze public statements or tweets. This case study has provided a set of tweets from Elon Musk along with stock data for TSLA which you may decide to use. Other text and financial data can be used and downloaded by yourself. You will use the VADER package in python to perform the sentiment analysis. You will then use your sentiment analysis along with time series analysis tools from multiple python packages to analyze relevant financial data. If using your own financial data, use a literature based approach to match the appropriate text data to financial data.

Deliverable Include:

- Github repository containing
    - README.md file that contains a brief description of the project and references utilized
    - DATA Folder containing raw data and cleaned and/or transformed data
    - SCRIPTS Folder containing all the python code for the analysis and model creation
    - OUTPUT Folder containing all files resulting form EDA and analysis
    - A one-page written summary of the results and model creation process

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

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 ▪ DATA Folder ▪ SCRIPTS Folder ▪ Written Document

	○ Use pdf format when possible and appropriate formatting ○ Use proper documentation for code
README.md	● Make a ## section explaining the contents of the repository and provide a short summary of what is included in the final deliverable ● A map of your documentation to show the organization of where each folder is located in the repository ● References in appropriate format included at the bottom. Cite all sources where information has helped the progress of the case study
DATA Folder	● All data that were used in exploration and model building should be located in this folder. ● Data that was transformed or cleaned should be contained in a subfolder ● The raw data provided from the case study should also be stored in a separate subfolder
SCRIPTS Folder	● This folder will contain all code pertaining to analysis ● All EDA scripts that helped to visualize and explore the data should be included. ● Create the model using python and execute all tasks involved in the case study ● Label each script according to the order they require to be executed in ● Label code cells with headers and use proper comments within code cells to ensure clarity
OUTPUT Folder	● This folder should contain all files and graphs that are generated from scripts ● Label each file according to each step that generated it
Written Summary	● A one page PDF document that explains how the EDA was conducted and the general process of model creation and validation