CS3 Rubric – Case Study Create

DS 4002 – Fall 2024 - Instructor: Loreto Alonzi
Due: TBD
Submission format: Upload a PDF and link to GitHub repository on Canvas

General Description: Submit to Canvas your PDF and link to the GitHub repository

Why am I doing this? This study is an opportunity to showcase your technical and conceptual skills in a unified project. You will explore the relationship between customer sentiment and product ratings using real-world tools like VADER Sentiment Analysis and hypothesis testing. This case mirrors the type of hands-on scenarios you may encounter in academic courses or professional environments.

What am I going to do? This assignment combines technical and analytical skills in a data science driven case study. You will produce a deliverable that demonstrates your ability to perform sentiment analysis and interpret its impact on product ratings. The deliverable will include include:

- Written portion PDF: Summarizing the project and findings, including a reference page.
- GitHub Repository: Containing code, necessary data, and supporting materials.

How will I know I have Succeeded? You will meet expectations on this case study by following the criteria outlined below.

Spec Category	Spec Details

Formatting	• Written Portion (1-2 pages)
	  ○ Submit the written portion as a PDF file.
	  ○ Include a reference page in IEEE citation style.
	• Data and Code
	  ○ Submit all code in a GitHub repository.
	  ○ Include any datasets used in the repository, as appropriate.
	  ○ The GitHub should be titled: CS-[insert first & last name].
	• References
	  ○ Include references on a separate page at the end of the written portion.
	  ○ Use IEEE citation style.

Written Portion	● Goal: The written portion should reflect your understanding of the study and its findings by addressing the following:
	● Problem Statement (1 paragraph):
	  ○ Summarize the problem of analyzing customer sentiment and its impact on product ratings.
	  ○ Discuss the significance of understanding customer sentiment in the e-commerce industry.
	● Plan of Action (1 paragraph):

	○ Provide a brief description of your plan to meet the deliverable. ○ Include a simple graphic outlining your analysis plan (e.g., a flowchart showing data collection, sentiment scoring, hypothesis testing, and conclusions). ● Results (1-2 paragraphs): ○ Discuss your findings clearly, including: ▪ Correlation between sentiment scores and ratings. ▪ Common words associated with high and low ratings. ○ Explain the significance of these findings in the context of improving customer satisfaction and product design. ● Reflection (2 paragraphs): ○ Reflect on the challenges faced during the study and how you overcame them. ○ Discuss what you could have done differently and how you plan to improve in the future.
Code	● Goal: Your code should include: ● Data Preparation and Preprocessing: ○ Clean and preprocess review data to prepare it for sentiment analysis. ▪ Download the data located in the "Data" folder on GitHub called "original_data.csv" ● Sentiment Analysis: ○ Perform sentiment scoring on the provided dataset using VADER Sentiment Analysis. ● Hypothesis Testing: ○ Test whether sentiment scores correlate with overall product ratings using statistical methods. ● Word Frequency Analysis: ○ Identify and visualize the most common words associated with high and low ratings. ● Comments and Documentation ○ Include comments in the code to make it easy for others to follow your process.
References	● At the end of the written portion, include a list of references in IEEE citation style. Ensure to cite all sources used that were not provided as part of the given materials.

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