CS3 Rubric - Charlottesville Restaurant Review Sentiment Analysis
DS 4002

Due: Refer to Instructor Rubric
Submission format: Upload conclusion PDF and link to GitHub repository to canvas

Why am I doing this? This case is an opportunity to apply data science skills learned in the class to a real-world application that is relevant and fun. Hopefully, you will expand your data science knowledge and start to think about where you want to take your skills in the future.

What am I going to do? In this case, you will combine various data science techniques to research Charlottesville restaurants, scrape websites, explore the data, and apply a sentiment analysis. You will create a deliverable that shows key insights and provides recommendations on how to best use reviews. The deliverable will include:

- GitHub repository containing data, scripts, and outputs
- PDF document containing key findings and references

Tips for success:

- Find introductory articles and rubrics in the reference materials fold in the GitHub.
- Use the data and code provided in the data folder in the GitHub.
- Take the time to get to know the data before starting.
- Find one meaningful result, it's okay if not everything is reproduced.
- This project is supposed to be fun and experiential!

How will I know I have succeeded? You will meet the expectations of the case study when you follow the criteria in the rubric below.

Spec Category	Spec Details

Formatting	• Repository: GitHub repo containing all materials. Title the repo as "CS-[first name-last-name]".
	  ○ Submit a link to the repo
	  ○ Contents in correct format
	  ■ Rubric (PDF)
	  ■ Data (CSV)
	  ■ Scripts (ipynb)
	  ■ Outputs (JPG)
	  ■ Conclusion document (PDF)
	• References
	  ○ All references not included in the hook should be listed at the end of the conclusion document
	  ○ Use IEEE Documentation style (link)

Conclusion File	**Goal:** One page documenting highlighting your process and most important findings from the case study. ● Summarize the case study (2-3 sentences) ● Document the analysis plan and key steps before beginning the case (2-3 sentences) ● Reflect on unknowns, bias and tricky analysis decisions (2-3 sentences) ● Highlight key finding (1 sentence) ● Discuss possible improvements and next steps (2-3 sentences) ● Reflect on the process of reproducing a case study, what were the successes and challenges? (2 sentences)
Code	**Goal:** produce the following scripts. ● Data: data set name ● EDA: potential questions to answer ○ What is the average rating for each restaurant? ○ How does the number of reviews change based on the day of week? Month? Year? ○ Relationship between sentiment and price of restaurant? ○ Relationship between sentiment and age of restaurant? ○ How does sentiment vary by time of year? ● VADER sentiment analysis: what is the sentiment in each restaurant's review? ● Logistic Regression: predicting average rating using average sentiment
Output	**Goal:** visualize results in a meaningful way. ● Include all outputs from the code ● Not all graphs and charts have to be included in the conclusion document

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