

## 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 folder in the GitHub.
- Use the data and code provided in their respective folders 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	<ul style="list-style-type: none"><li>• Repository: GitHub repo containing all materials. Title the repo as "CS-[first name-last-name]".<ul style="list-style-type: none"><li>○ Submit a link to the repo</li><li>○ Contents in correct format<ul style="list-style-type: none"><li>■ Rubric (PDF)</li><li>■ Data (CSV)</li><li>■ Scripts (ipynb)</li><li>■ Outputs (JPG)</li><li>■ Conclusion document (PDF)</li></ul></li></ul></li><li>• References<ul style="list-style-type: none"><li>○ All references not included in the hook should be listed at the end of the conclusion document</li><li>○ Use IEEE Documentation style (<a href="#">link</a>)</li></ul></li></ul>

Conclusion File	<p><b>Goal:</b> One page documenting highlighting your process and most important findings from the case study.</p> <ul style="list-style-type: none"> <li>● Summarize the case study (2-3 sentences)</li> <li>● Document the analysis plan and key steps before beginning the case (2-3 sentences)</li> <li>● Reflect on unknowns, bias and tricky analysis decisions (2-3 sentences)</li> <li>● Highlight key finding (1 sentence)</li> <li>● Discuss possible improvements and next steps (2-3 sentences)</li> <li>● Reflect on the process of reproducing a case study, what were the successes and challenges? (2 sentences)</li> </ul>
Code	<p><b>Goal:</b> produce the following information.</p> <ul style="list-style-type: none"> <li>● Data: data set name</li> <li>● EDA: potential questions to answer <ul style="list-style-type: none"> <li>○ What is the average rating for each restaurant?</li> <li>○ How does the number of reviews change based on the day of week? Month? Year?</li> <li>○ Relationship between sentiment and price of restaurant?</li> <li>○ Relationship between sentiment and age of restaurant?</li> <li>○ How does sentiment vary by time of year?</li> </ul> </li> <li>● VADER sentiment analysis: what is the sentiment in each restaurant's review?</li> <li>● Logistic Regression: predicting average rating using average sentiment</li> <li>● Word cloud: what words are associated with positive and negative reviews?</li> </ul>
Output	<p><b>Goal:</b> visualize results in a meaningful way.</p> <ul style="list-style-type: none"> <li>● Include all outputs from the code</li> <li>● Not all graphs and charts have to be included in the conclusion document</li> </ul>

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