Restaurant Review Case Study Rubric

DS 4002 – Alice Vadney
Due: TBD
Submission format: Upload link to GitHub repository

Why am I doing this?
This case study is an opportunity to leverage your data science skills by using sentiment analysis to compare quantitative and qualitative reviews for restaurants. This assignment will allow you to engage with a familiar topic (restaurant Yelp reviews) in a new way with the addition of sentiment analysis and regression.

What am I going to do?
The GitHub repository for this case study can be found at https://github.com/alicevadney/CS3. You will obtain a dataset of Yelp reviews for a selection of Charlottesville restaurants, along with a sentiment analysis of the written review. After obtaining the data, you will perform an ordinary least squares regression to determine the relationship between the sentiment of a review and its quantitative "star" rating. You will then provide a scatterplot of the results. The deliverable will include:

- A data dictionary
- Well-documented, commented source code
- All outputs (graph, regression results)
- GitHub repository containing all materials

Tips for success:
- Make your file names and your variable names easily interpretable for the grader/viewer. It should be clear from the name what exactly is contained in the file.
- You will be working in Python. Familiarize yourself with the language to the best of your ability before starting this case study in order to complete it more efficiently.

How will I know I have succeeded?
You will meet expectations on this case study when you successfully follow and complete the criteria in the rubric below:

Spec Category	Spec Details
Formatting	● One GitHub repository (submitted via link on Canvas) ○ Create a new repository named 'CS3_Restaurants' that contains the following: ■ README.md ■ LICENSE.md ■ Your source code ■ OUTPUTS: containing your visual outputs ■ The dataset
README.md	● Goal: this file should orient everyone who views the repository. ● Brief summary of what you've produced for the case study. This does not have to be very detailed but should provide enough information for viewers to understand.
Source Code File	Well-documented Jupyter Notebook file and that contains the code used to execute your analysis. You must include: ● Simple EDA for the data ● Frequency table for each restaurant ● An OLS regression and output table ● A scatterplot of the values with a line representing the regression ● Comments throughout explaining the code and interpreting output
References	● All references should be listed at the end of the document ● Use IEEE Documentation style

Acknowledgements: Thank you to Professor Alonzi for providing the rubric structure!