Food.com Case Study Rubric

DS 4002 – Spring 2024 - Bella Holloman
Due: May 10
Submission format:

- **Upload link to github repo to canvas AND turn in hard copy**

Individual Assignment

Why am I doing this?
This is your opportunity to complete a case study on the length of recipes and how that correlates with the reviews of the recipe. As you work, you will learn how to perform VADER sentiment analysis and discover results.

What am I going to do?
The Github repository can be found at https://github.com/bellaholloman/CS3-DS4002 . You will utilize the data and code to recreate the case to find out the correlation between length of recipe and its reviews. You will perform a VADER sentiment analysis in Python on Food.com recipe reviews. There are so many ratings in this dataset that you will choose a random 1,000 to perform the analysis on. The recipe ratings are also in the dataset to check the accuracy of the sentiment review. There are other materials in the repository to help you understand why and how to do this.

Deliverables include:

- Dataset of Food.com recipes
- Data Appendix
- Correlation numbers
- Source Code
- Complete Github Repository

Tips for success:

- Familiarize yourself with VADER in Python before beginning
  - There is material in the repository to read
- Make sure you have Python and all of the correct libraries downloaded
- Make sure the code is opened and the data is downloaded with the correct name

How will I know I have Succeeded?
You will meet expectations on CS3 Create Case Study when you follow the criteria in the rubric below.

Formatting	o One GitHub repository named “CS3-yourname” o This repository should contain these folders/files: o README.md o LICENSE.md

	○ Data Folder ○ Original large dataset ○ Dataset of 1,000 random chosen ratings you performed the analysis on ○ Code Folder ○ All source code used with comments ○ References.md
README.md	● Goal: Summarize the case study you have performed ○ This text file should overview what you did for the case study and what your results are ○ For the results, include the correlation coefficient found
LICENSE.md	● Goal: Have a license for your repository ○ Generally the MIT one
Data Folder	● Goal: Include datasets used for the case study ○ The original large Food.com dataset given ○ The dataset of the 1,000 random recipes chosen for the case study
Code Folder	● Goal: Include all code used for the case study ○ Any code used for the entire case study ○ Comments throughout to show understanding of the code
References.md	● Goal: Include all references ○ A text file of all references used to complete the case study, whether it was a material given or found separately