Case Study Rubric

Due: TBD
Submission Format: Provide a link to your GitHub repository on Canvas

General Description:
Submit the link to your GitHub repository on Canvas. This case study provides an opportunity to showcase your technical, analytical, and programming skills by analyzing Yelp restaurant reviews. Your mission is to extract actionable insights from review data, identify trends and keywords that influence restaurant ratings, and offer recommendations for improving customer satisfaction. This hands-on project simulates real-world data science scenarios, allowing you to apply sentiment analysis, keyword extraction, and visualization techniques to real-world data.

Why am I doing this?
This case study gives you the chance to demonstrate your abilities to extract valuable insights from textual data, similar to challenges faced in industry. By leveraging natural language processing (NLP) and sentiment analysis, you will uncover trends in Yelp reviews and provide actionable recommendations for improving customer satisfaction. You will combine statistical methods, programming skills, and visualization techniques to analyze customer feedback and its impact on restaurant success. This project simulates work that could be encountered in the restaurant industry, marketing, or consumer behavior research.

Tips for Success:

- **Organize your repository** clearly to make it easy for others to navigate and understand your work.
- **Plan your analysis steps** before diving into coding. A clear roadmap will help you stay focused.
- **Leverage visualizations** to help communicate your findings clearly and effectively.
- **Write clear, well-commented code** to ensure that others can follow your process and replicate your results.
- **Test your models thoroughly** to ensure that your analysis is robust and your results are reliable.
- **Highlight key insights in your report**—don't just present the data, explain why it matters.

- **Manage your time effectively**—start early to avoid last-minute pressure.

GitHub Repository Requirements:
Your GitHub repository, titled "Yelp_Rating_Analysis", must contain the following:

1. **README.md:**
    - A brief summary introducing your case study project, including the research question and the methods used.
    - A high-level summary of your findings and any key insights or recommendations.
    - Basic instructions for running the code and replicating the results.
2. **LICENSE.md:**
    - Specify the licensing for your project (e.g., MIT, GPL).
3. **Source Code File:**
    - A well-documented Python script (or Jupyter notebook) that implements the sentiment analysis and keyword extraction process.
    - The script should include the following steps:
        - Preprocessing the Yelp review text (e.g., removing stop words, lemmatization).
        - Performing sentiment analysis on the review text.
        - Identifying key phrases and keywords associated with higher or lower ratings.
        - Visualizing sentiment trends over time or across different restaurant attributes.
        - Providing clear comments explaining each step of the analysis.
4. **Data Folder:**

- Include all datasets used in the analysis. Clearly name and organize the files, and include a description of each dataset.

5. **REFERENCES.md:**

   - A markdown file citing all references in IEEE style.

   - Brief annotations explaining how each source informed or contributed to your analysis.

How Will I Know I Have Succeeded?
You will meet expectations on this case study when you follow the criteria outlined below:

Spec Catagory	Spec Details
Formatting	● **Title:** "Yelp_Rating_Analysis" ● **Folder Structure:** ○ **README.md:** A concise project overview, including your research question, methodology, key findings, and instructions for running the code. ○ **LICENSE.md:** Licensing information for your project. ○ **Source Code File:** Python script or Jupyter notebook. ○ **Data Folder:** Include all data used in the analysis, with clear file naming. ○ **REFERENCES.md:** A markdown file citing all references in IEEE style with annotations.

README.md	The README file should: ● Introduce the purpose and structure of your project. ● Summarize the methods you used and key findings. ● Provide instructions for running the code and replicating your results.
Source Code File	Your source code file should be well-documented and contain: ● Preprocessing steps (e.g., removing irrelevant words, tokenization). ● Sentiment analysis using logistic regression or a similar model. ● Extraction of keywords or phrases associated with higher or lower ratings. ● Visualizations (e.g., bar charts, word clouds, sentiment over time) that illustrate the relationship between review content and ratings. ● Clear comments explaining each step of your analysis, particularly how you interpret results from sentiment analysis and keyword extraction.
REFERENCES.md	This file should include: ● All cited sources, formatted in IEEE style. ● Brief annotations explaining how each source contributed to your analysis.

Evaluation Criteria:

- **Completeness:** Inclusion of all components (repository, code, analysis, report).
- **Clarity:** Clear documentation and visualizations that effectively communicate your analysis.
- **Insight:** Depth and relevance of your analysis and how well you identify key factors influencing ratings.
- **Accuracy:** Proper application of sentiment analysis and keyword extraction techniques.