**Case Study Rubric**
Due: TBD
Submission format: upload pdf and link to github repo to Canvas

**Why am I doing this?** This study is an opportunity to showcase your technical as well as conceptual skills in a unified project. The case study will represent a hands-on scenario similar to one you may be presented in a professional environment.

**What am I going to do?** At this point, you have likely accumulated a variety of technical and conceptual skills in data science. You will now have the opportunity to combine and express these skills in an independently-driven case study. You will ultimately provide a deliverable that covers all requirements, including significant results and conclusions. The deliverable will include:

- Conclusions PDF - along with references
- GitHub repository - containing code and necessary data

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

| **Category** | **Details** |
|---|---|
| Formatting | Submit each component listed in the rest of this rubric as advised below.<br>- Conclusions Document<br>  - Submit your conclusions as a PDF file.<br>- Data & Code<br>  - Submit code created for all portions in a GitHub repository.<br>  - Include any additional data that was used in GitHub if necessary.<br>- References<br>  - References should be included on a separate page at the end of the conclusions PDF file.<br>  - IEEE citation style |
| Conclusions Document | Discuss your thought process in the case study and any conclusions made following the format below:<br>- **Introduction** (1 paragraph): summarize the problem presented in the study as well as its importance.<br>- **Analysis Plan** (1 paragraph): Discuss what your plan was to meet the demands of the deliverable. Include the methods of analysis used and some background info on them.<br>- **Results and Conclusions** (1-2 paragraphs): Discuss the results of your study in complete sentences. Answer the questions: Is price level and review sentiment correlated? Is there a specific price level of restaurant that the restaurant group should open, and if so, what price level do you recommend? |

| | |
|---|---|
| | - **Reflection** (1 paragraph): Discuss the challenges you met while completing this study and how you overcame them. Also discuss what you could have done differently and how you plan to improve in the future. |
| Code & Data | Include your code and data used in your GitHub repository for the following tasks:<br>- Exploratory Data Analysis<br>  - Explore the distribution of restaurant price level, the number of reviews per restaurant, and anything else that seems interesting.<br>  - Use the cleaned datasets philly_restaurants.csv and philly_reviews.csv from the DATA folder on GitHub.<br>- VADER Sentiment Analysis<br>  - Use the VADER sentiment analysis package to assign numerical sentiment scores values to each Yelp review, ranging from -1 (negative) to 1 (positive) with 0 being neutral. Then take the mean sentiment score across all reviews for each restaurant.<br>- Correlation Analysis<br>  - Create a graphic that shows the average review sentiment across restaurant price levels.<br>  - Calculate a correlation coefficient and assess its statistical significance, Spearman's rank coefficient is recommended but you can use any correlational analysis method that is appropriate for this situation.<br>Be sure to include comments so a viewer can understand your process. |
| References | At the end of the conclusions document please include a list of any outside references you used beyond the given reference materials. (IEEE citation style) |

Acknowledgements: Special thanks to Professor Rasero for the example rubric formatting. This structure is pulled from Streifer & Palmer (2020).