Case Study: Connotation Revenue Rubic
Due: TBD
Submission format: Upload link to GitHub repository via Canvas

Individual Assignment

General Description: Submit to Canvas a link to your case study repository including the takeaways and findings from your work

Preparatory Assignments: Previous data science experiences and the resources provided in this document.

Why am I doing this? This case study is a great way to combine all the skills of a data scientist into one project and showcase what you're capable of. Whether it be finetuning already existent skills or learning new ones, you'll undoubtedly gain value in both your hard and soft skills. By applying your skills in a real-world application, you'll gain relevant job experience of working through problems and coming to conclusions. Furthermore, the text-based date will either build you a new and valuable skill or sharpen your already existent sentiment analysis abilities.

- Course Learning Objective: Analyze text-based data
- Course Learning Objective: Practice the data science lifecycle
- Course Learning Objective: Interpret model finding into understandable conclusions

What am I going to do? First read the supporting documentation to understand the landscape of the issue. Decide on the scope of the project and determine what data will be needed to pursue this thoroughly. Once you collect the relevant data don't be hasty to jump into model building. Identify what model will prove most beneficial to you and proceed with model creation and interpret the result in context to answer the question at hand. Does headline connotation correlate with revenue generation?

- Necessary code and result documentation

Tips for success:

- **Don't jump into this blind.** Think about what data you'll need and whether you can reasonably collect it. There are many models out there, weigh the pros and cons of each before deciding on one.
- **There does not necessarily need to be a right answer.** This is the nature of building models. There are many uncertainties, justifying your decisions is the important part, not necessarily the outcome.

How will I know I have Succeeded? You will meet the expectation outline in the following rubric.

Spec Category	Spec Details

Formatting	• Github Repository containing all the necessary materials
	• Contents include:
	  ○ README.md
	  ○ LICENSE.md
	  ○ SRC folder
	  ○ DATA folder
	  ○ FIGURES folder

README.md	• Goal: An overview of the repository. Guide someone who has never interacted with this project before
	• Use easy to understand language
	• SRC section
	  ○ Explain how to install the code as well as use it
	• DATA section
	  ○ Create a data dictionary outlining the important aspects of the data
	  ○ Include your actual data here as well
	• FIGURES section
	  ○ Any images created throughout the process
	• REFERENCES section
	  ○ Any additional references used to be successful

LICENSE.md	• Goal: This file explains to a visitor the terms under which they may use and cite your repository. • Usually, the MIT license is appropriate.
SRC folder	• Goal: This folder contains all the source code for your project. • Include all of the code created in the process • The high level documentation for the code exists in the README.md but include and supplemental documentation you feel necessary
DATA folder	• Goal: This folder contains all of the data for this project • Place all your data here. If your data does not fit in GitHub, outline how to access it
Figures folder	• Goal: This folder contains all of the figures generated by your project • Include all the figures created. This includes charts, graphs, and images for example. • Also include any relevant background information to understand the figures
References	• All references should be listed at the end of the document • Use IEEE Documentation style
Findings	• Detail the relevant finding in one document. Rember that the intricacies of the model are not important, the findings and reasoning are. Convey your finding so that they make sense to nontechnical people as well.