

Case Study Rubric: “The Office”

DS 4002 – Spring 2024

Due: May 10

Submission format: Upload your code and images into a github repository.

Individual Assignment

General Description: Submit to canvas a link to your github repository.

Why am I doing this? As a data scientist, you will often come across complicated, real-world questions like the one presented in this case study. You will have to figure out how to tackle these issues and how to present your results. This study gives an example of one way to tackle a social science question, using sentiment analysis, and asks you to create visuals that communicate your results. You will gain practice to use going forward in your data science career while answering a difficult question.

What am I going to do? You will read all materials provided for this case study (Rubric, Hook Document, Helpful Files). These are available in the physical folder as well as on the provided github. You will then download the starter code in the github and open it in R. You will run the provided code which loads useful packages and the dataset. Then, you will use the VADER package to conduct a sentiment analysis on the dataset. After conducting your analysis, you will analyze your results and compare the representation of different genders in “The Office”. You will create informative and interesting graphs to present your results. Finally, you will create a presentation that summarizes your analysis for a general audience.

Deliverables include:

- Edited starter code file that includes your VADER analysis.
- Graphs that visualize your results
- A presentation of your analysis

All of this will be submitted electronically via a link to a github repository built for this case study.

Tips for success:

- Be creative. Make interesting and colorful graphs for your presentation. This will interest your audience and add to the success of your project.
- Learn a lot. Don’t hesitate to ask your instructors questions. The goal here is to learn as much as possible and grow your data science skills.
- Have fun!

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

Formatting	<ul style="list-style-type: none"> Repository - a github repo including the following materials and folders <ul style="list-style-type: none"> README.md: a general explanation of the project and the materials included in this repo Folders: <ul style="list-style-type: none"> Data: the dataset should be in this folder Code: all your code should be in this folder Images: all graphs created for this project should be in this folder Presentation: your presentation should be in this folder
README.md	<ul style="list-style-type: none"> Goal: This introduces your project Include the following sections <ul style="list-style-type: none"> Introduction Motivation/Goal Included in this repo Directions for how to do this project
Data	<ul style="list-style-type: none"> Goal: Include the data you used in this project Add your dataset to this folder in a csv format
Code	<ul style="list-style-type: none"> Goal: Include all the code you used to complete your project and create graphs Should be an edited version of the starter code Should include commented explanations and clear steps. A person reading it should be able to follow what's going on.
Images	<ul style="list-style-type: none"> Goal: Visualize and communicate your results Make informative and interesting graphs Should be colorful and draw in audience Should explain what you've found in your sentiment analysis Make 2-3 different graphs
Presentation	<ul style="list-style-type: none"> Goal: Communication your analysis to a general audience Include the following sections <ul style="list-style-type: none"> Introduction Motivation/Goal Steps Important Results (including your graphs) Next steps, future analysis Sources

Acknowledgements: Special thanks to Jess Taggart from UVA CTE for coaching on making this rubric. This structure is pulled from [Streifer & Palmer \(2020\)](#).