CS – PUBLISHING A BEST SELLER Rubric

DS 4002 – Spring 2023

Due: TBD

Submission Format: Link to github repository (canvas assignments)

Individual Assignment

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

Preparatory Assignments – Class sessions about case study reading; Lessons in analysis; Lessons in data visualization.

Why am I doing this? As a data scientist, it is important to learn how to research broad topics and produce a result. Through your experience as a student of data science, you have developed the skills needed to perform analysis and interpret results. What this project will do is allow you to use those skills in a valuable lesson. You will take the skills you have honed and answer a large research question that will require finding data, assessing the data, and using it to answer the bigger question at hand, finally producing a result that will be understandable by non-data-scientists. This project will make you a better data scientist and make your skills more applicable in the real world.

What am I going to do? You will begin by reading the one page Hook document that explains the scope of the case study and the deliverable you are meant to produce. You will then examine the provided materials, and potentially find some of your own, to begin working with the dataset. You will write code in your chosen vice and complete analyses on the data. Upon interpreting your results, you will present them in a slideshow.

Tips for success:

- Try to remain unbiased and examine all potential paths with an open mind
- Try different analysis techniques: linear regression may work to find correlation but try multiple methods of analysis for the best results
- Try to think about the context of the data such as what goes on during the time of year when a book is published or what the genres are of the most popular books or how many people are rating the books. This is a very difficult to predict topic so try not to get discouraged

How will I know I have succeeded?

You will produce a deliverable that points to specific factors that are the most important in publishing a successful book with analysis and data to back up your claims. Your deliverable will meet the criteria detailed in the rubric below.

Spec Category	**Spec Details**

Formatting	• Repository – a github repository containing all materials
	  ○ Submit a link to the repository
	  ○ Everything is contained in the repo or linked to it if appropriate
	  ○ Contents
	  ▪ README.md
	  ▪ DATA folder
	  ▪ Figures folder
	  ▪ SRC folder
	  ▪ Presentation folder
	  ▪ LICENSE

README.md	• Goal: This file will explain what can be found in the github repo • Use level 2 headings to indicate what folders can be found in the github and regular text to explain them • DATA ○ Create a data dictionary using the table format in github • Include references and acknowledgements at the end
DATA folder	• Goal: Provide the data used for your analysis such that someone may replicate your results • Upload a CSV file containing the data • Include the source of the data in the description of the CSV file
Figures Folder	• Goal: Show figures used in data analysis and interpretation so that a viewer may understand your conclusions • Include all figures created in your code • Write a description of each uploaded figure
SRC Folder	• Goal: Provide code you wrote for the analysis and cleaning so that your results may be replicated and understood • Upload a file containing the code written in your chosen programming language

Presentation Folder	• Goal: Present your results
	• This folder should contain a PDF of your slides presenting your findings
	• Slides
	  ○ Title slide
	  ▪ Name
	  ▪ Title
	  ▪ Date
	  ○ Agenda Slide
	  ○ Motivation Slide
	  ▪ Include research question
	  ○ Data Slide
	  ▪ Data dictionary
	  ○ Modeling Approach and Analysis Plan Slide
	  ○ Results Slide
	  ○ References and Acknowledgements

LICENSE	• Goal: Explain to the readers the terms in which they may use and share your work • The MIT license is the default recommendation

Acknowledgements: Thank you to Professor Alonzi for the guidance in creating this rubric and the assignment overall. Thank you to Kathryn Iapalucci and Olivia Kaczmarskyj for their help and contributions this semester in our group projects.