Sephora Lip Product Case Study Rubric

DS 4002 - Spring 2025 - Yiyang Bu
Due: TBA
Submission format: Upload GitHub repository link (submit to Canvas)

Why am I doing this? This case study is meant to introduce you to the basics of image analysis. It will give you the chance to practice using data analytical techniques on real datasets. Throughout the process, you will be exposed to modeling approaches and exploration opportunities to get hands-on experiences with image analysis. It will also provide you with the chance to document your analysis professionally.

What am I going to do? The Github repository for this case study can be found here. You will obtain the Image URLs and product details in the dataset provided. In Github, you will find an example script on how to transfer the image URLs to image formats, which can be used in the analysis. After collecting all of the product images, you are encouraged to explore the dataset, including understanding the key variables and creating visualizations. For the actual analysis part, you will build a logistic regression and a convolutional neural network (CNN) model to predict the correlation between lip product colors/visual patterns and product popularity. You can find additional guidance in the Github script. Whenever you encounter obstacles, you could refer to the materials in the "Technical" folder for more concept and model explanations.

Tips for success:

- Research the different components of creating a CNN to understand what produces a quality model.
- Don't get stressed! Learning to create regression and CNN models on image data can get overwhelming.
    - Utilize the provided resources, research online or take a step back to help you understand the problems.

Final Deliverable should include:

- Output graphs to display the accuracy and loss of your models
- A data dictionary
- Well-documented script
- A Github repository storing all materials used and produced, with a README.md documenting your analysis process

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

Spec Category	Spec Details
Formatting	● One Github Repository (submitted via link on Canvas) ● Title the Github Repository "Sephora Lip Product Case

	Study_YourName” ● The repository should contain the following folders: ○ A README.md file ○ A SCRIPTS folder ○ A DATA folder ○ A OUTPUTS folder ○ A REFERENCES folder
README.md	● Goal: This file serves as an orientation to everyone who accesses your repository. It should enable them to understand this project without further explanations. ● Provide a brief summary of the purpose of the case study, the script you created, and your findings from the models you conducted
SCRIPTS folder	● Goal: This folder contains all the source code for your project. ○ Make sure to clearly state what the script was used for in the script name and header. ● Add detailed comments to the script so that someone who never worked on the script before could understand what is happening.
DATA folder	● Goal: This folder contains all of the data related to this project. ● Include the dataset that was used to train your train, if changes were made to the existing dataset. ○ Clearly label if the data within the folder is part of the initial dataset or if the data is the result of code. ● The final data dictionary will also be added in this folder.
OUTPUTS folder	● Goal: This folder contains all of the outputs created by your project. ● Include any outputs (graphs, tables, figures) generated by your models and EDA. ● Use informative names for your files.
REFERENCES folder	● Goal: Cite additional resources related to this case study project. ● Any additional resources that were used to help you create your model should be included here. ○ Cite sources using IEEE Documentation style.

Acknowledgements: Special thanks to Professor Alonzi for the structure of this rubric!