Pinterest and Color Theory Case Study Rubric

DS 4002 – Fall 2024 – Ashley Rommel

Due: December 9

Submission Format: Upload Link to GitHub repository on UVA Canvas

Individual Assignment

General Description: Submit to Canvas a link to your GitHub repository for this case study.

Why am I doing this?

The goal of this assignment is to reproduce the results of a previous DS 4002 project. Reproducibility is an important part of a data science project because it ensures reliability and validity of the project. This minimizes the risk of errors or biases and can help strengthen the evidence supporting the hypothesis. It is good practice for data scientists to make their projects reproducible and for other data scientists to be able to reproduce data. Also, this case study will grow your knowledge of image analysis techniques and models to look at the distinguishing factors of warm and cool tones.

What am I going to do?

The GitHub repository for this case study can be found at https://github.com/ashleyr05/Pinterest-and-Color-Theory-Case-Study. You will be provided with two datasets of warm and cool tone images from Pinterest's recommendations. With these datasets you will use Python to determine if there are distinguishing factors in warm and cool tone images. You will do this by analyzing and creating a Random Forest Classifier for the YUV values of the images. After creating the model, you will create a Feature Importance and PCA plot to analyze if there is a difference between warm and cool tone images.

Your final deliverables should include:

- The provided datasets for the images from Pinterest
- A data dictionary
- Well documented, commented source code
- A GitHub repository containing all materials used

Tips for success:

- Make your file names and variable names easily interpretable
- Get comfortable with Python and the techniques that will be used in this case study
- Work to understand the code and your model, so that you know what your findings are and what it means

How will I know I have succeeded?

You will know you have succeeded when you successfully follow and complete the criteria in the rubric below:

Spec Category	Spec Details

Formatting	• One GitHub repository (submitted via link on Canvas)
	  ○ A README.md file (which auto displays)
	  ○ A LICENSE.md file (use MIT at default)
	  ○ Source Code File
	  ○ The provided data for warm and cool tone images
	  ○ A REFERENCES.md file

README.md	• Use markdown headers to divide content.
	• Make an H2 (##) section explaining the contents of the repository
	• Section 1: Summary of what you've produced for the case study to orient people to your repository
	• Section 2: Software and platform section
	  ○ The type(s) of software you used for the project.
	  ○ The names of any add-on packages that need to be installed with the software.
	  ○ The platform (e.g., Windows, Mac, or Linux) you used.
	• Section 3: A Map of your documentation.
	  ○ In this section, you should provide an outline or tree illustrating the hierarchy of folders and subfolders contained in your Project Folder, and listing the files stored in each folder or subfolder.

Source Code File	Well commented scripts used for the image and data analysis for this case study. In the source code, you must include: • Box plots for both warm and cool tone datasets • Histograms for both warm and cool tone datasets • Random Forest Classifier Model for data • Confidence Scores for model • Feature Importance and PCA plots
REFERENCES.md	Markdown File citing any resources referenced in helping you executing and interpreting this case study in IEEE Documentation style.

Acknowledgements: Thank you to Professor Alonzi for providing the rubric structure!