Case Study Rubric – MET Museum Image Classification

DS 4002 – Spring 2025 - Instructors: Javier Rasero; TA: Mercedes Mora-Figueroa

Submission format: Upload link to github repo to canvas. Complete individually.

Why am I doing this? This case study gives you an opportunity to apply your skills in image analysis so you can organize and classify image data for an art archive. Through the assignment, you will be exposed to the thought-process of practicing data science in the real world, as well as understand how to use data science skills to encourage public appreciation in the arts and culture.

What am I going to do? You will use the MET Museum API to obtain image data and image metadata from the MET museum archive. Using this data, you are expected to develop a classification model that is suitable for use in a shared public archive. The GitHub repo contains sample code of how to do API requests and store into a dataframe, however you are strongly encouraged not to copy directly from the provided code. After cleaning your dataset as necessarily, you will develop a model that is trained on the dataset. Your final product should evaluate the success of the model and the parameters used to develop it. Deliverables include:

- A final extracted dataset from the API used for your image analysis
- A data dictionary
- Github repository and data set – to provide resources like code and data
- Well documented, commented source code

Tips for success:

- Explore the API requests! Try not to build your code directly from the code provided, but rather use it as a foundation. You can either take data from one department in the MET museum, or from a specific year, author, etc. It depends on the model you want to develop.
- Ask questions! Reference materials are provided, but do not hesitate to reach out to your instructors to help guide you through the process. Part of being a data scientist is communicating your ideas and take in feedback from others.
- Start early! There can be unexpected hiccups down the road. You will be less stressed if you pace yourself

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

Formatting	● Create a GitHub repository that includes o **Title " [first name] Image Classification Case Study"** o **Informative names for all your files in the repo**

	o **Contents:** ▪ README.md ▪ LICENSE.md (use the MIT license) ▪ DATA FOLDER ● Final API data file (images used) ● Data Appendix as a PDF, which will include text, figures, and descriptive statistics. See guidelines here ▪ SCRIPTS folder ● Coding scripts made + used ▪ OUTPUT folder ● Images of the output of your project
README.md	● Goal: This file serves as an orientation to everyone who comes to your repository, it should enable them to get their bearings. ● Use markdown headers to divide content. ● Make an H2 (##) section explaining the contents of the repository ● Section 1: 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 2: 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. ● Section 3: Instructions for reproducing your results. In this section, you should give explicit step-by-step instructions to reproduce the Results of your study. These instructions should be written in straightforward plain English, but they must be concise, but detailed and precise enough, to make it possible for an interested user to reproduce your results without much difficulty.
Rubric	● Goal: Explain the purpose, task, and criteria. This is where the details live to answer questions for the student. ● Use this document's format. There is no need to reinvent the wheel. ● The goal is about 2 pages.
References	● All references should be listed at the end of the document ● Use IEEE Documentation style (link)

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