DSC540 Final Project

The source I chose for the Final Project was the API from the Harvard Art Museum (<https://github.com/harvardartmuseums/api-docs>). The API has several variables available to collect about many works of art in their collection. There is a limitation about how many entries can be pulled so my project performed a GET request of 99 instances. I was able to pull 17 different variables with the API request and used a keyword of “dog” as a part of my search to pull specific objects with “dog” in the title.

Once I received the data, I then created a For Loop to pull the items into a dictionary that then transformed the data into a list. I was then able to convert the data to a Pandas Dataframe so that it would be in a formatted dataset.

I performed two types of data transformations. There was a variable called “Date of Last Page View” and I used the textbook as a guide and the “datetime” library to reformat the month, date, and year. I decided to add the new format as a new column so that both formats could be available for analysis. For the second transformation, I focused on the variable Technique ID. The reason I picked this one is because the numbers had “.0” at the end and this was not necessary and could potentially cause confusion in future analysis. I first removed all entries that had NaN listed as the value and then converted the values to a string to remove the extra decimal value. I then converted the values back to integers and they were formatted the way I planned on them being.

Finally, I wanted to make sure I could verify the data was in a formatted dataset ready for processing. All variables are available in a Pandas Dataframe but I also exported as a CSV file for additional verification.

One issue I ran into with this API was with missing keys in the JSON data. In this particular API, every key was not available for every object and it was unclear where the distinction was drawn by the API maintainers. A challenge for myself going forward is how I could write an IF/ELSE statement that accounted for any missing keys in an API GET request.