Problem Statement:

The problem statement has not changed: The goal of this project is to create a model that classifies stellar bodies (galaxies, stars, etc) given their spectral characteristics values. This model is theoretically possible since each class of stellar body emits different spectral values, creating a numerical relationship.

Data Processing:

The dataset used remains the same: "Stellar Classification Dataset - SDSS17". No changes concerning the dataset were made. While the dataset contains over 14 different attributes corresponding to the name, label, location, etc. of the stellar bodies, we are only interested in the attributes of the 5 color filter values (**u**, **g**, **r**, **i**, **z**) and the class, since only the spectral filter values play a role in helping to determine the stellar class. As such, a new .csv file was created that only contains the filter values and the class. This is all the pre-processing that needs to be done, as the data is numerical.

Machine learning model:

I had initially chosen the Bayes Classifier model as seen in Assignment 2, as I thought that estimating the probability that a body belongs to a class can be computed using the probabilities of the values. However, I found difficulty in the matter and I'm now looking for different models to use. The other two options I am considering for the model are "Multinomial Logistic Regression model" and probabilistic models such as "MAP" or "MLE". I will contact the TPMs in the few upcoming days to determine the correct model to use.

Next steps:

Seeing that my model still hasn't been settled, I'll firstly reach the TPMs to figure the issue at hand. I'll then regularly update this page.