The questions here are based on the paper "Deep-learning-based regression model and hyperspectral imaging for rapid detection of nitrogen concentration in oilseed rape (Brassica napus L.) leaf". The relevant data ("Meanspectra.csv").

* 'N Values' - label
* 'band1' to 'band512' - features
* 'Dataset'- binary variable.

We wish to model 'N Values' as a function of 'band1' to 'band512'.

1. Perform exploratory data analysis and present your results and insights. We expect to see summary statistics, informative visualizations, and anything else you find relevant for summarizing high dimensional data.
2. Let be samples of 'N Values' for 'Dataset'=0 and let be samples of 'N Values' for 'Dataset'=1. The two samples were collected in different experiments. Examine whether the two samples are similar or not.

Use 'Dataset' to split the dataset into two sets: train-set ('Dataset' =1) and test-set ('Dataset'=0). We wish to model label as a function of the features.

1. Fit the appropriate statistical model for prediction of 'N Values' (not necessarily one of the models suggested in the paper).
   1. Explain why you chose this type of model.
   2. Perform model performance analysis and explain in analysis in details.