Final Project: Due 07/17

To incorporate what you have learned this summer, your final project is to do a small Machine Learning project using Python. Here are the first steps:

* Pick a dataset from the UCI Machine Learning Repository (<https://tinyurl.com/ya78oakh>) that has a default task of either classification or regression. The dataset needs to have more than 100 instances and more than 3 attributes. For those newer to Machine Learning I recommend trying to also have less than 250 instances and 10 attributes. However, feel free to pick any dataset that meets the minimum requirement.
* Select a classification or regression algorithm to use to solve your problem. For those newer to Machine Learning I recommend trying Linear Regression, Logistic Regression, Decision Trees, or Support Vector Machines. For those who have Machine Learning experience feel free to pick a harder algorithm or a set of easier algorithms.

Once you have selected your dataset and algorithm, write a report similar to my example data science problem. Here are the requirements for the report:

* Provide roughly 2 paragraphs of explanation on how your selected method works. I recommend 1 paragraph giving a high-level overview and 1 paragraph giving technical details on the algorithm.
* Perform exploratory data analysis on your chosen dataset. This includes:
  + At least 2 visualizations of your data
  + Explaining what the variables in your dataset mean
  + Identifying if there is any missing data
* Perform preprocessing on your dataset. This includes:
  + Strategy for replacing or fixing missing data (if applicable)
  + Creating the training and testing set
  + Scaling data if applicable for your selected algorithm
* Apply your selected machine learning algorithm to your dataset
  + Fit the model using scikit-learn and report the training error (MSE for Regression, Accuracy for Classification)
  + Make predictions for your test set and report the testing error (MSE for Regression, Accuracy for Classification)
* Perform model selection
  + Test 2 other models and see if the performance increases or decreases. Note: Using the same algorithm with different input features is a different model.
* Provide a 1 paragraph conclusion on your major findings.
* The entire project should be done in a Jupyter notebook. Become familiar with Markdown so you can annotate your code to give explanation where needed. For those familiar with Latex, you can do Latex commands in Markdown cells in Jupyter Notebooks. Once completed, export your Jupyter Notebook as a pdf.

There are no right answers to your specific project. I am looking for your thought process as you work through the project. Make sure to comment your code and explain why you made the choices you made. This is meant to be a learning exercise for you, so the more you explain throughout your code, the more I can help you at the end. I want all of you to succeed, so if you have any questions just let me know. I will be happy to help and clarify any requirements if needed.