Assignment #1

Formatting:

       If you're coding in a Jupyter Notebook, just submit your ipynb file with all the outputs. If you're using a py file, include all the outputs in a pdf along with your Python code.

From scikit learn toy data set:

https://scikit-learn.org/stable/datasets/toy_dataset.html

1- Read wine dataset.
2- Print and review the 'DESCR' of this dataset.
3- Plot histogram of all 13 wine features ('data')
4- Use pandas.DataFrame.describe to produce basic descriptive analysis (such as mean, min, max, std) for all features in a table. For percentiles get 10, 25, 50, 75, 90 percentiles.
5- Are there any null values in any of features?
6- Split the data set into test and train (20 vs 80 percent). Build a random forest classifier using:

    from sklearn.ensemble import RandomForestClassifier
    clf = RandomForestClassifier()
    And then get the accuracy of the classifier for the test set.

Grading Criteria:

Total: 100

    1- 10
    2- 10
    3- 25
    4- 10
    5- 5
    6- 40