Homework #1

due Wednesday, Sep 4 (end of class)

individual assignment

In this assignment we are practicing working with python, pandas and numpy on an automotive dataset. You will need to submit a Jupyter notebook with the code that answers the questions in this assignment via Canvas.

Download the following file from Canvas:

- auto-mpg.csv

Create a notebook cell for each of the following questions:

1. [5pts] Create a pandas data frame using the data in the ‘auto-mpg.csv’ file.
2. [5pts] Display the first 7 rows in the data frame.
3. [5pts] Display the number of entries and the data types of all the fields.
4. [10pts] Display the unique values of the ‘cylinders’ field.
5. [5pts] Show the mean, std, min, and max of 'mpg' attribute.
6. [5pts] Plot the histogram of 'mpg' field.
7. [10pts] Create a new data frame that groups the entries by ‘cylinders’ and aggregates them by taking the mean for all the fields. Display the new data frame.
8. [5pts] Display the index of the new data frame.
9. [10pts] Create a new field ‘disp_per_cyl’ that gives the displacement per cylinder.
10. [5pts] Create a bar plot for ‘mpg’ vs ‘cylinders’.
11. [5pts] Create a line plot for ‘disp per cyl’ vs ‘cylinders’.
12. [5pts] Save the new data frame into a file name 'summary.csv'.
13. [10pts] Create and display a numpy matrix using only the ‘mpg’ and ‘disp_per_cyl’ fields.
14. [5pts] Display the transpose of the new numpy matrix.
15. [10pts] Create and display a copy of the third column of the transposed matrix.