Python Review Homework

Python libraries (Pandas & Numpy)

Problem 1:

- Load the boston dataset by using Pandas read_csv().
- Remove column zero (the tag for this column is 'Unnamed: 0'
- Remove column tagged as 'dist' and join the two parts of the dataframe (to the left and right of the column 'dis') back together in a new dataframe called df2.
- Calculate the mean of column called 'age' and add it as a new column with the mean value repeated for all rows.

Problem 2:

- Generate a vector of 1000 random numbers between 0 to 100 taken from a normal distribution.
- Plot a histogram of these numbers with number of bins equal to 10.
- Calculate the average of these numbers by using numpy method mean().
- Plot a line with a red color from the mean point on the histogram plot in y direction with a width of 3 to show the location of mean in the histogram plot.
- Given the following matrix, multiply this matrix with its transpose (matrix multiplication) then calculate the dot product of the first matrix with its transpose. Compare the results of the two operations. Elaborate on the results and explain if this result can be generalized (you may want to try a couple of other examples and compare the results to generalize).

$$\begin{matrix} 3 & 6 \\ 4 & 9 \\ 1 & 5 \end{matrix}$$

- Generate a vector of 5 random integer values between 1 and 10. Multiply the vector by its transpose. What is the relationship of the resultant matrix and the original matrix?