Home Work #4

1. You will make use (only) of the numpy library to perform preliminary analyses of the auto mpg data in the data file auto-mpg.csv.

   - Place the data file auto-mpg.csv in your current directory.
   - Read it into an array using genfromtxt().
   - There are a few missing entries for horsepower (HP) which should show up as `nan`s. Use boolean indexing / masking to remove those elements.
     Print out the records you choose to delete.
   - Eliminate duplicate records.
     Print out the records you choose to delete.
   - Use cov() to compute the covariance of MPG against Cyl; use the returned 2x2 matrix to compute the correlation using

     $$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X * \sigma_Y}$$

     $$\text{where } \sigma_X^2 = Cov(X,X); \text{ and } \sigma_Y^2 = Cov(Y,Y).$$

   - Print the column name, covariance matrix and the correlation value.
   - Repeat for Displacement, HP, Wt, Accel, ModelYr, and Origin
   - Now compute the standard deviations of MPG and HP using std(). Use these numbers for σ_X, σ_Y in the formula for correlation. Does the value agree with your previously computed value? If not, explain.
   - You will submit one Python script file firstname.py which I should be able to execute by
     `python3 firstname.py`
     Note: the data filename will not be a parameter in this assignment.