Name: DATA 506 Assignment # 3 (50 Points)

We will use here a dataset D3.csv available on OAKS. There are many ways to import data in python. The pandas package seems to be the best one.

Q1: Let the first column of the data set be the explanatory variable x, and let the fourth column be the dependent variable y. [That is: ignore columns 2 and 3 for now]

(a) [10 points] Run simple linear regression to predict y from x. Report the linear model you found. Predict the value of y for new x values 0.3, for 0.5, and for 0.8.

(b) [30 points] On the same data, run polynomial regression for $p = 2, 3, 4$. Report polynomial models for each. With each of these models, predict the value of y for a new x values of 0.3, for 0.5, and for 0.8.
Cross-validate to choose the best model. Describe how you did this, and which data was used for what.

Q2: [10 points] Now let the first three columns of the data set be separate explanatory variables x1, x2, x3. Again let the fourth column be the dependent variable y. Run linear regression simultaneously using all three explanatory variables. Report the linear model you found. Predict the value of y for new (x1,x2,x3) value of (0.3, 0.4, 0.1).