ML Diagnostics

1)First I read the data from csv file

2)Then I view the correlation between the features and the prices

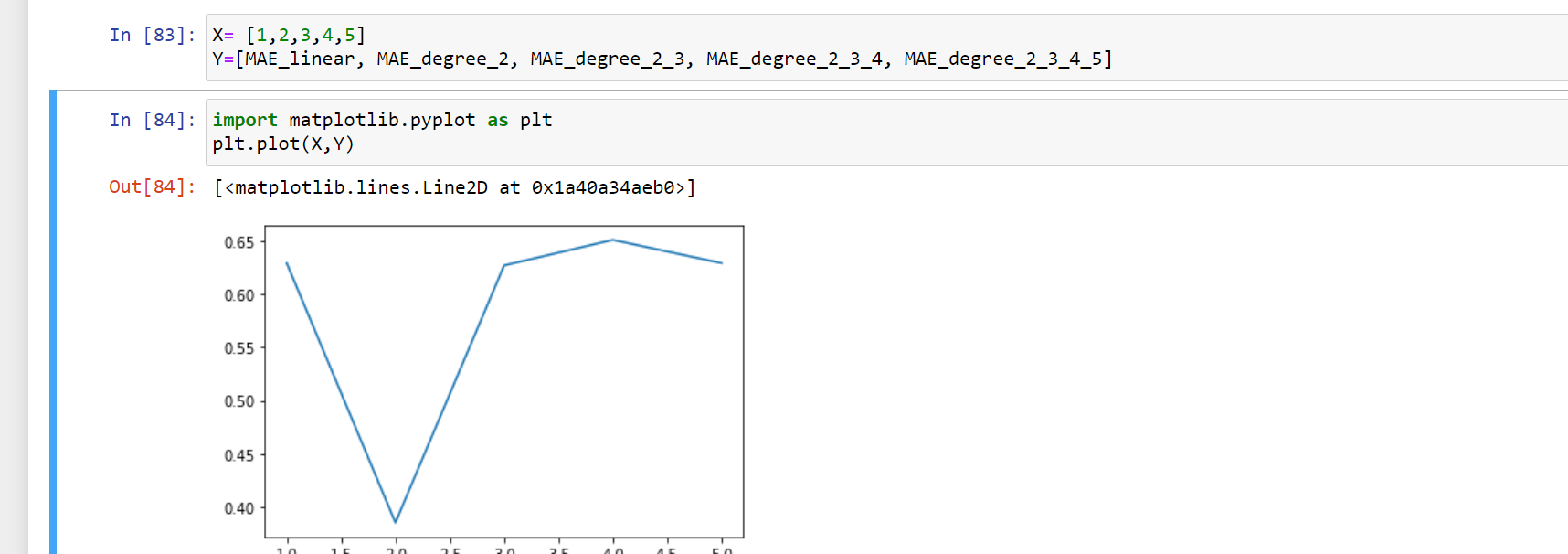
This is done by plotting each column with price to inspect for correlation

According to the analysis of correlation graphs I have concluded that the following features are the most influential

features = ['bedrooms','bathrooms','sqft\_living','floors','view','grade','sqft\_above','sqft\_basement','yr\_built','sqft\_living15','price’]

3) then I normalized the date

4)the I split the date 70% for training and 30% for testing

5) I implemented linear regression then polynomial regression of different degrees 

I plotted the mean square values of different polynomials with respect to the degree of polynomial in order to detect the best degree of polynomial to be used in hypotheses function

Best one is at degree 2