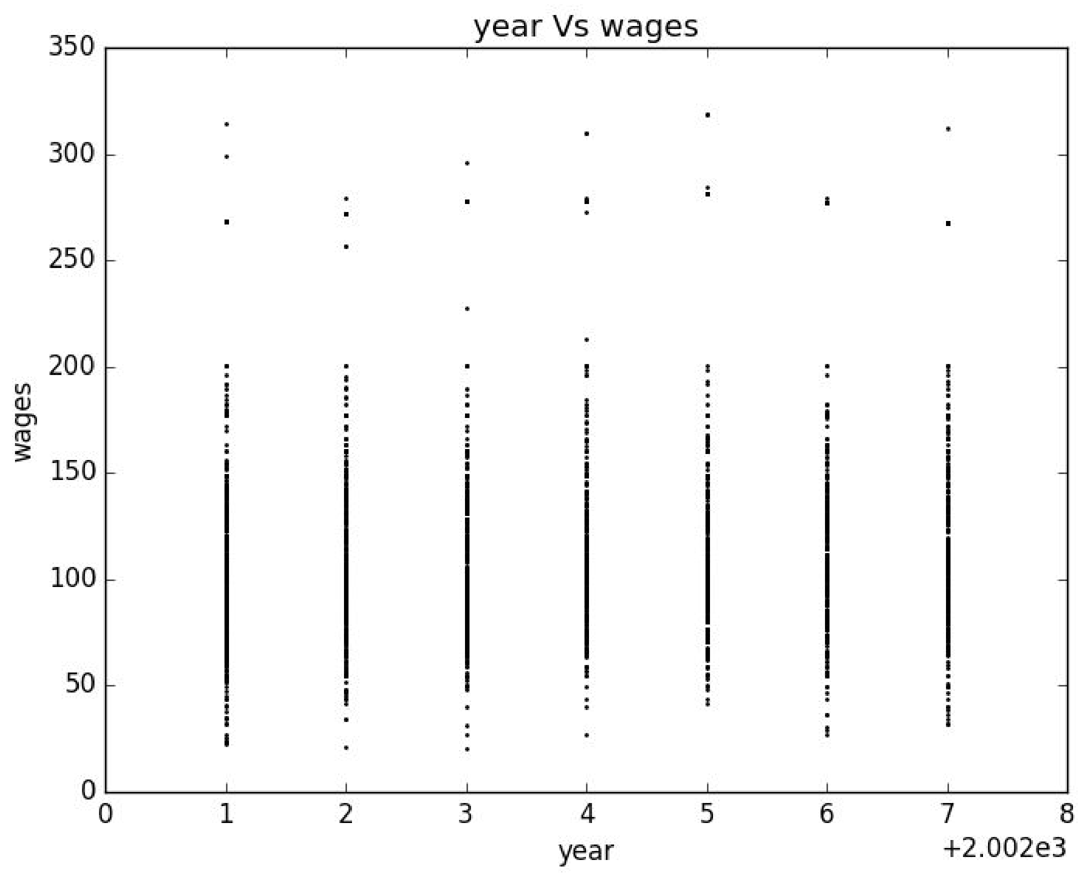
In problem 3, we are given with dataset of various parameters and wages. In this problem I try to fit various order polynomials using the following code. while running these program use python3 and it should contain modules like numpy , matplotlib etc.

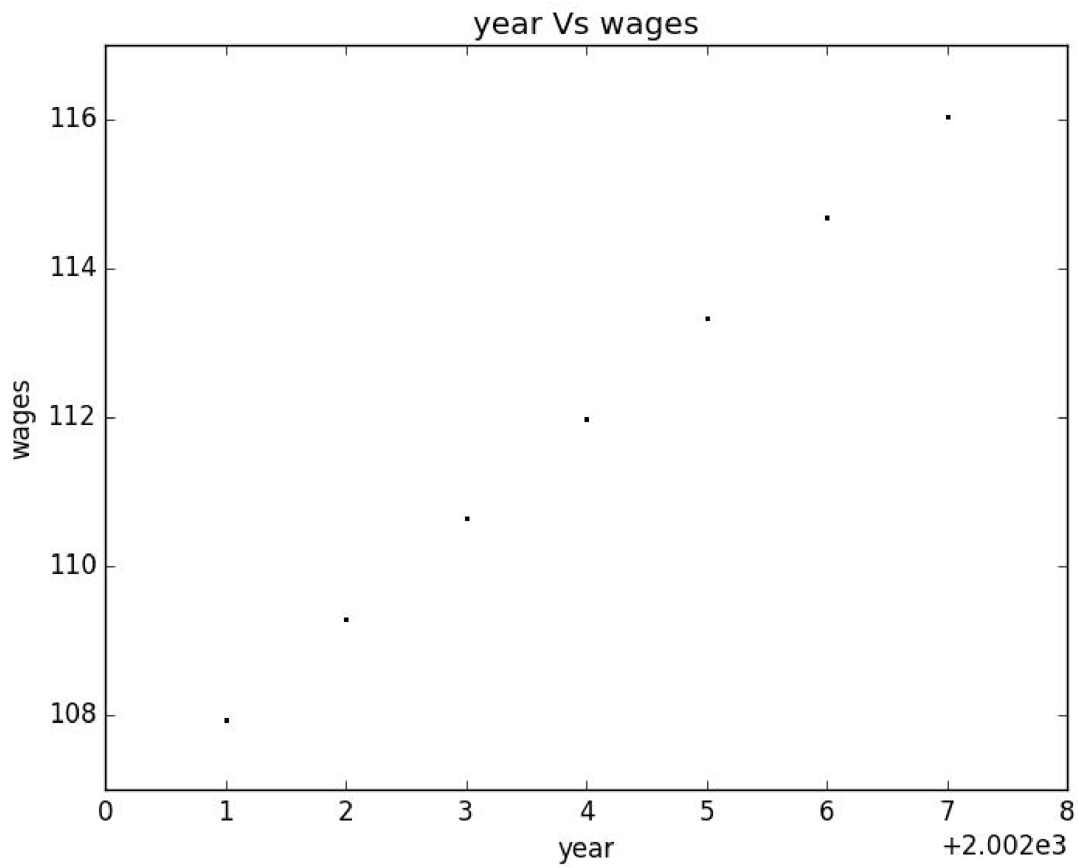
The code calculates the fitted polynomial and fitted values of wages w.r.t given parameters such as in part a) year , part b) age, part c) education. It also plot the fitted and given dataset using matplotlib module. It calculates the coefficient of polynomial using univariate polynomial regression in which vector **w** = inverse (**d**_transpose*__d__) * (**d**_transpose*__y__), where **d** is matrix and **y** is the given dependent variable. **w** is the column vector consisting coefficent of polynomial.

The function defined as regression in the code , gives **w** , **y_fitted** and mse(mean square error). For our first part i.e part a) we have **x =** year and **y** = wages. The curve of given data is :

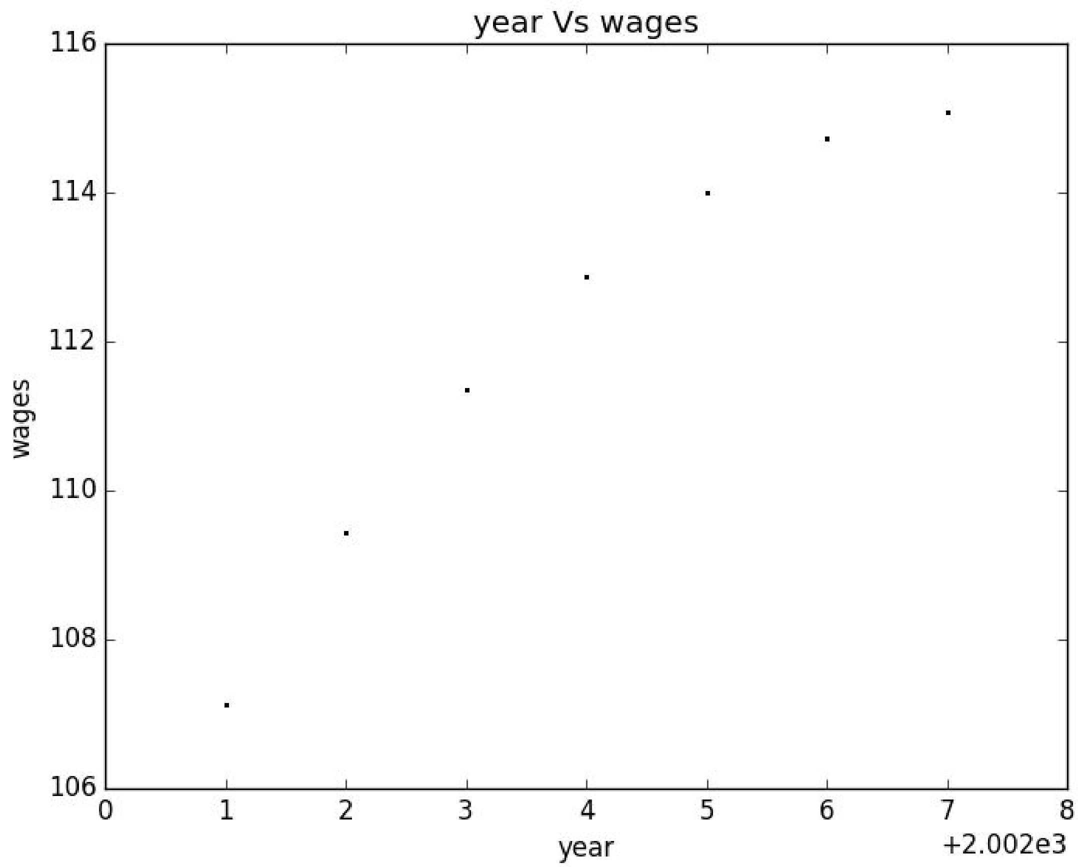


Let take various order polynomial to get the best possible fitting having least square error:

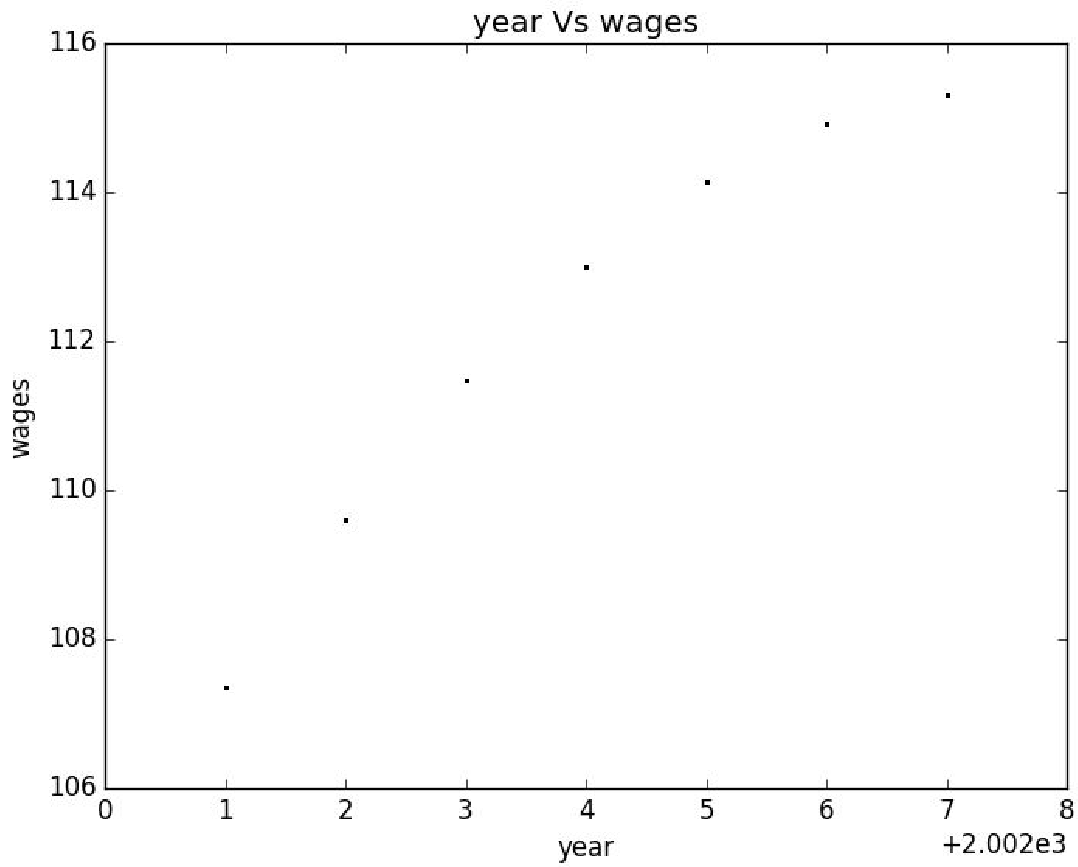
1) n = 1 , w vector : [-2.59586155e+03 1.34987402e+00]



2) for n = 2 w vector : [-7.92910125e+05 7.89326447e+02 -1.96411230e-01]

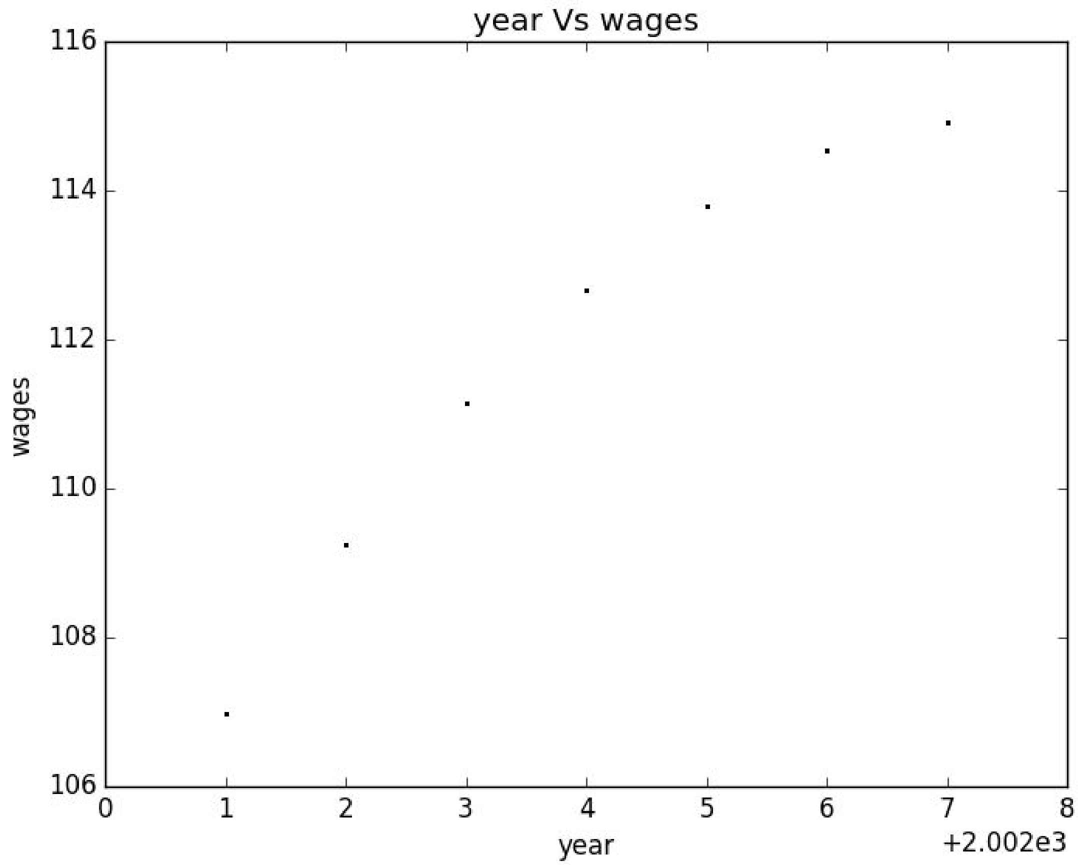


3) for n = 3   w vector : [-5.13368938e+05   3.95310852e+02 -1.19788349e-02
-2.86548784e-05]



year Vs wages
116
114
112
110
108
106
wages
0
1
2
3
4
5
6
7
8
year
+2.002e3

4) For n = 4   w vector : [-2.47478812e+05   7.53280640e+01   6.43234700e-02 2.06901314e-04   -1.10578391e-08]

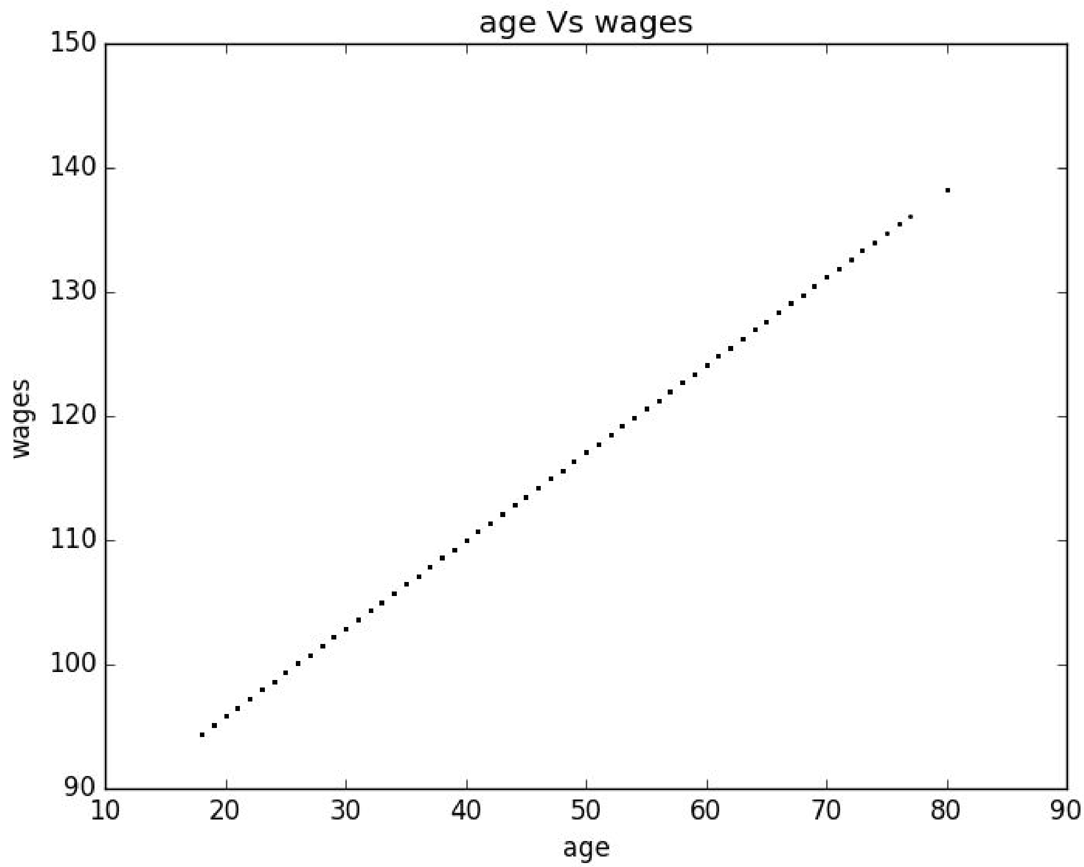


As we increase the order of polynomial the mse get increased . it is found that for n = 2 mse is the least so for wages vs year polynomial of order 2 is the best choice.

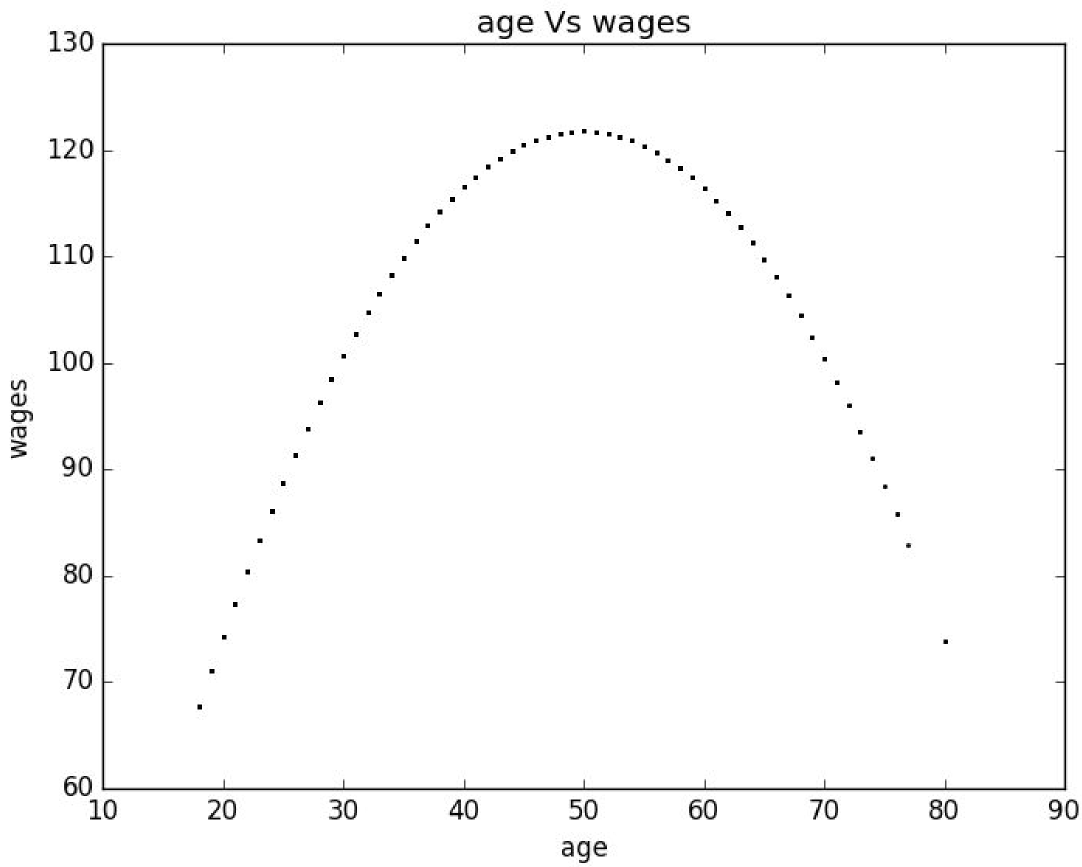
Part b) we have **x =** age and **y** = wages. The curve of given data is :

Let take various order polynomial to get the best possible fitting having least square error:

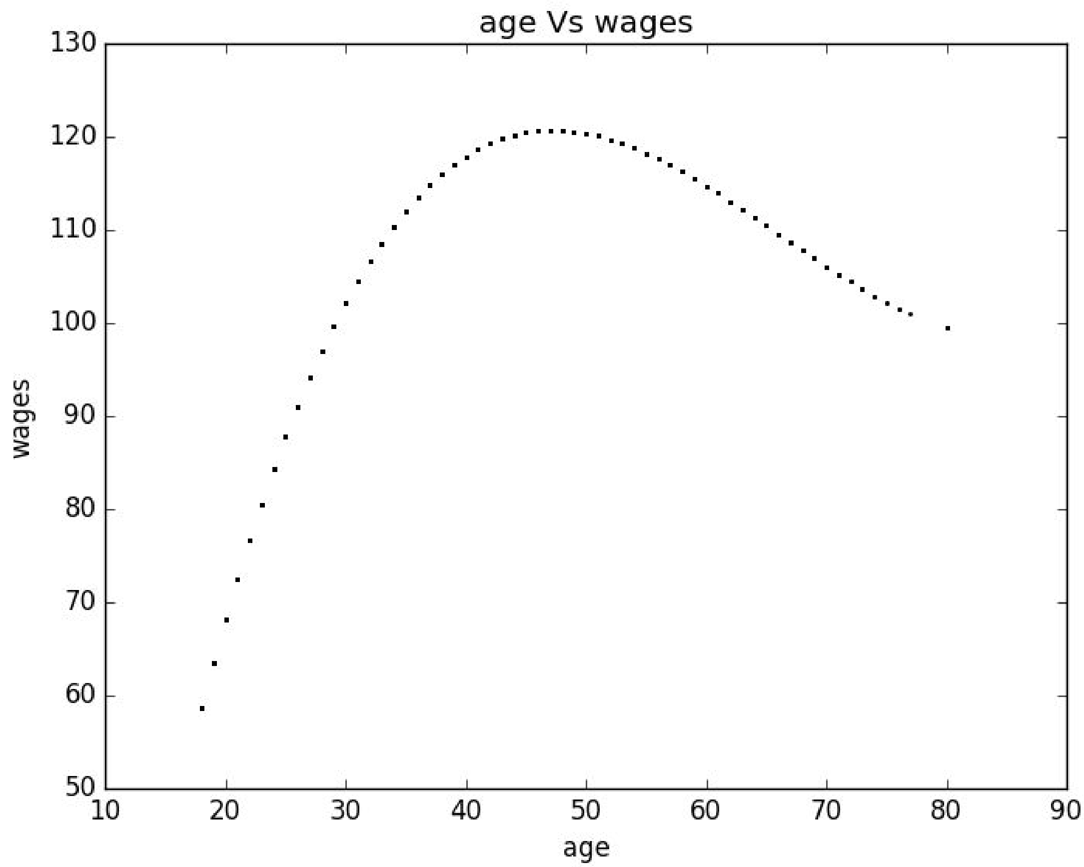
1) n = 1 , w vector : [81.70473544 0.70727593]



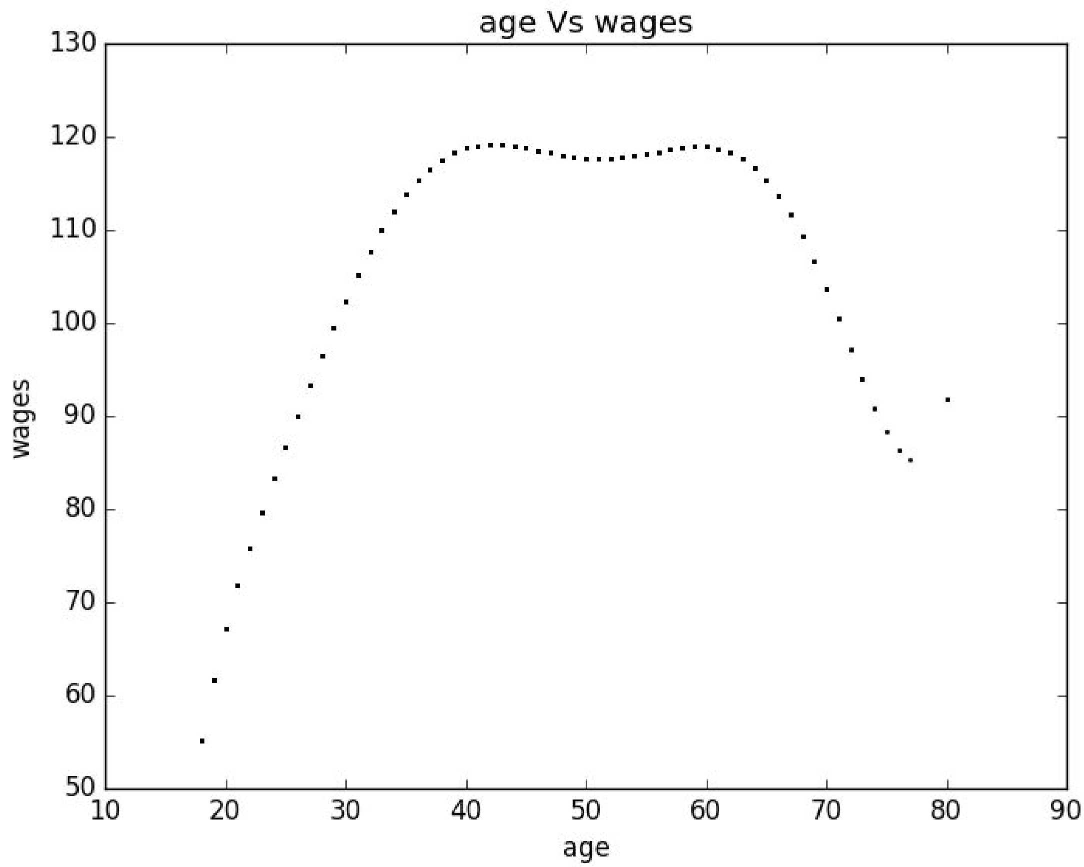
2) for n =2   w vector : [-10.42522426   5.29403003 -0.05300507]



3) for n = 3 w vector : [-7.52439142e+01   1.01899915e+01 -1.68028587e-01
8.49452197e-04]

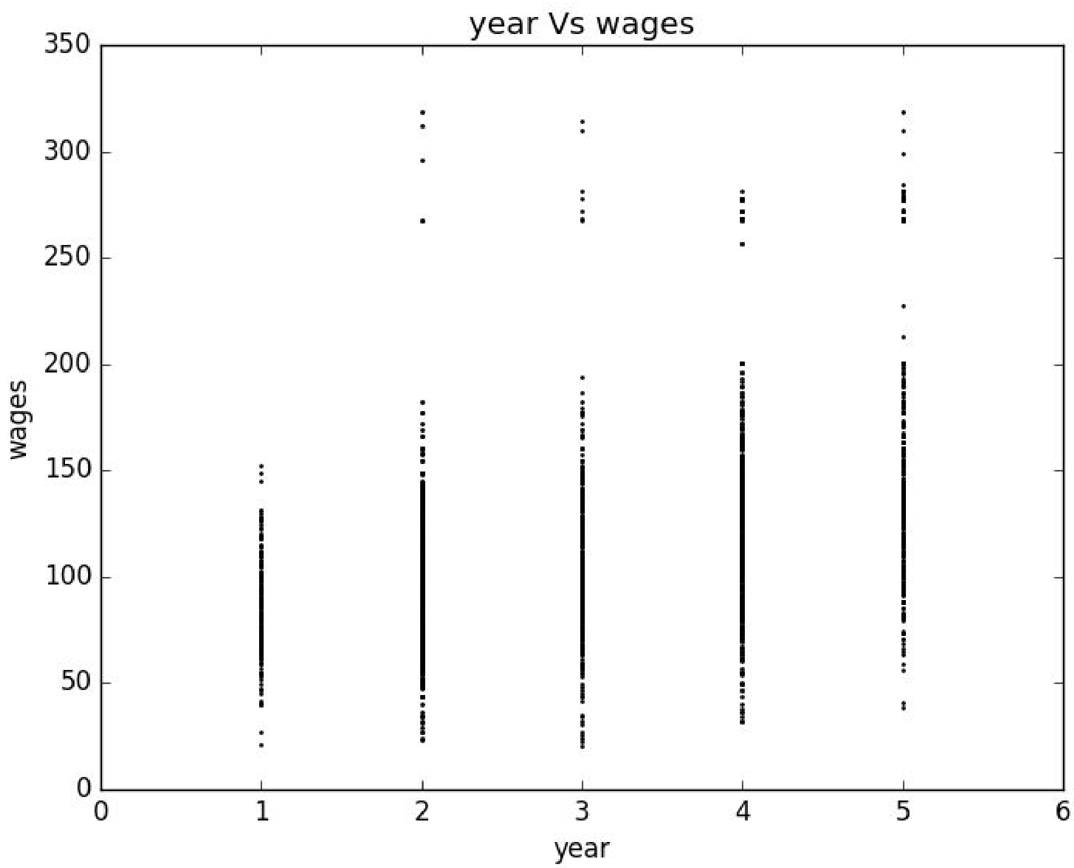


4) for n = 8 w vector : [-1.72007479e+03 3.48936274e+02 -2.93124333e+01 1.36410031e+00 -3.78381577e-02 6.37483019e-04 -6.36255094e-06 3.43637841e-08 -7.67305010e-11]

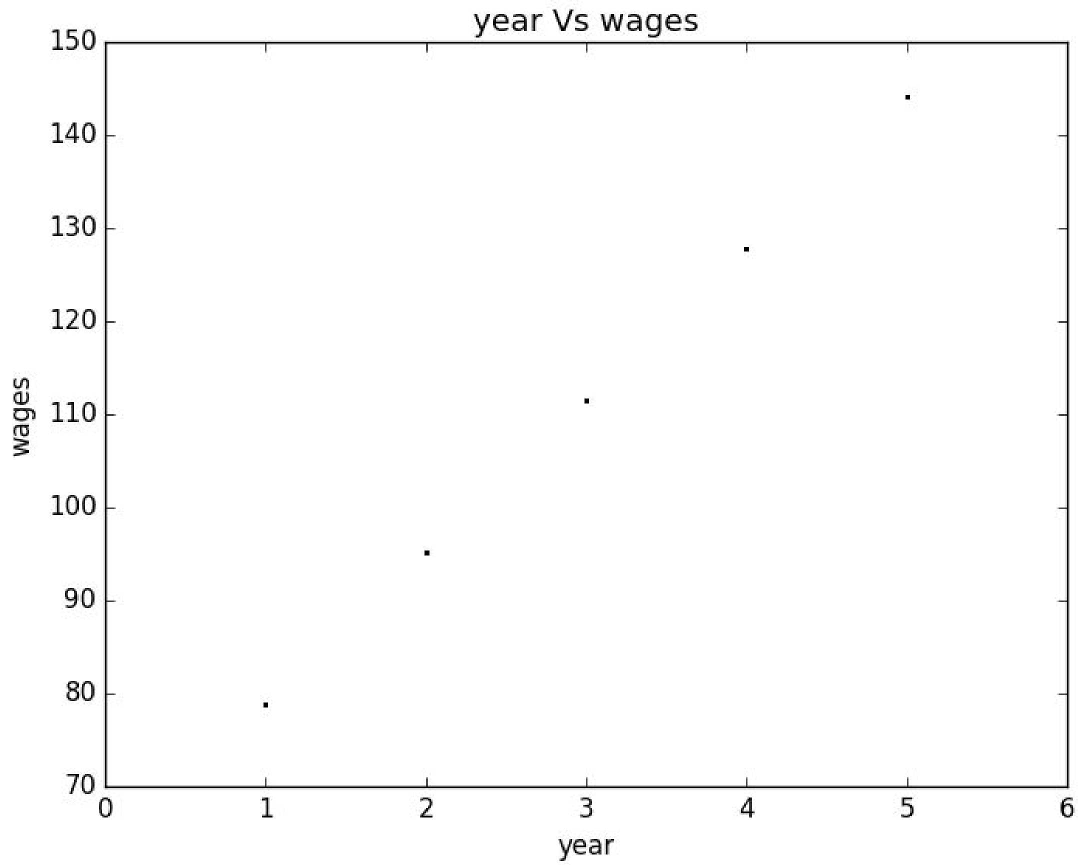


As we increase the order of polynomial initially till n =8 the mse get decreased and then increase . it is found that for n = 8 mse is the least so for wages vs year polynomial of order 8 is the best choice.

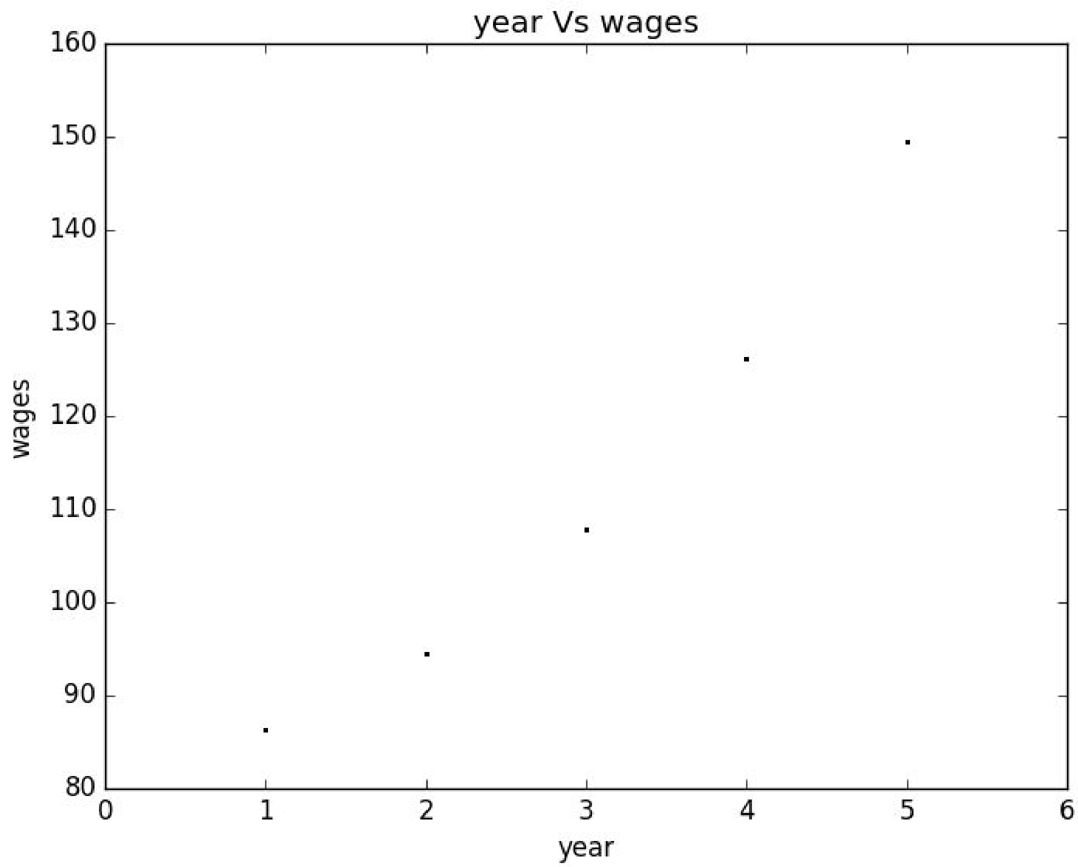
Part c) we have **x =** education and **y** = wages. The curve of given data is :



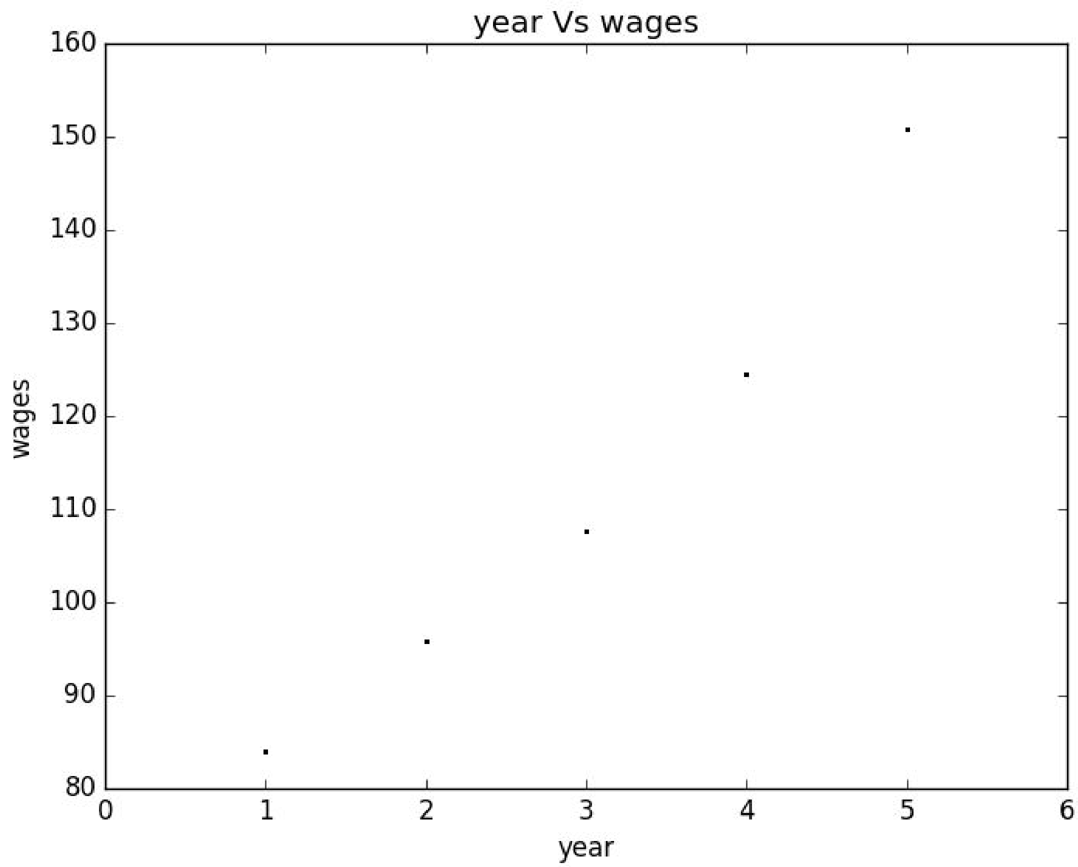
1)For n = 1   w vector : [62.54440413 16.33196149]



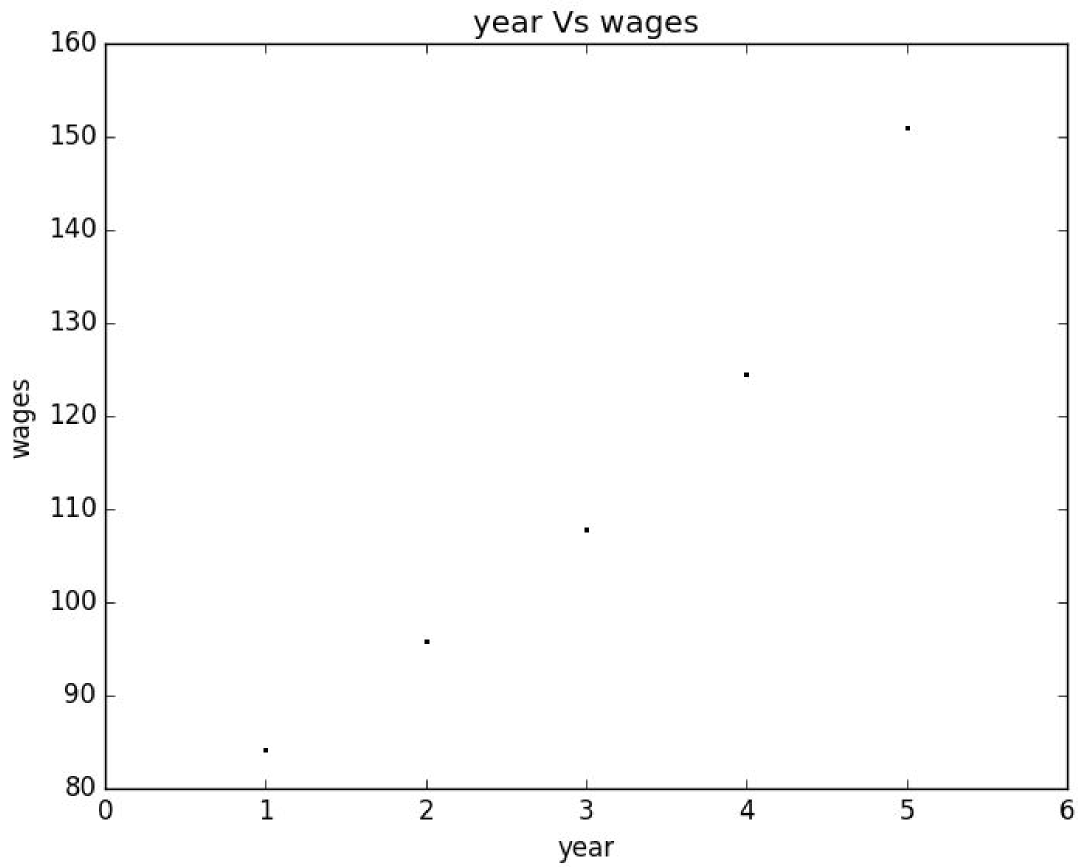
2)For n = 2 w vector : [83.21303888   0.5946616   2.53390844]



3) n =3 w vector : [67.73388956 20.23449693 -4.68263027   0.79248555]



4) n = 4   w vector : [6.90224599e+01   1.78379257e+01 -3.22398882e+00   4.38411791e-01
  2.96062952e-02]



As we increase the order of polynomial initially till n =4 the mse get decreased and then increase . it is found that for n = 4 mse is the least so for wages vs year polynomial of order 4 is the best choice.

In these problem since we have plotted the wages dependency on year , age and education ; but the finding shows that due to very high mean square error mse we can not get very good prediction from the above regression.