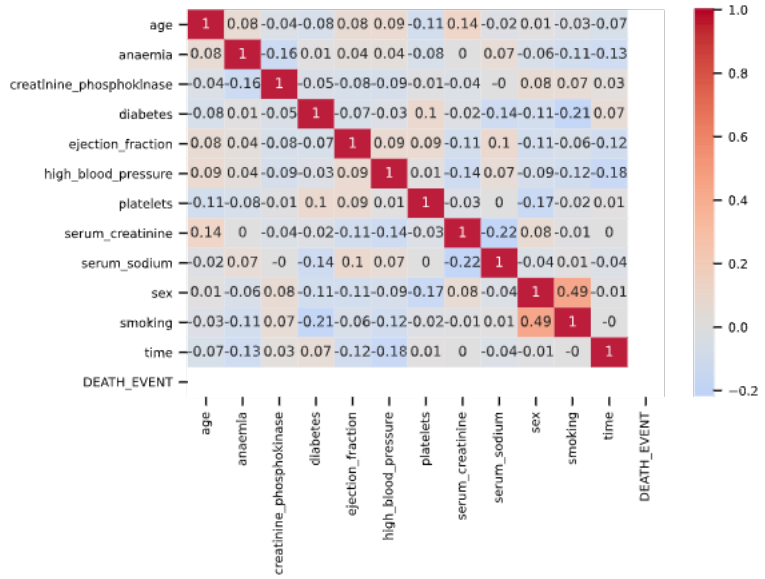
Yubin Ye U53651145

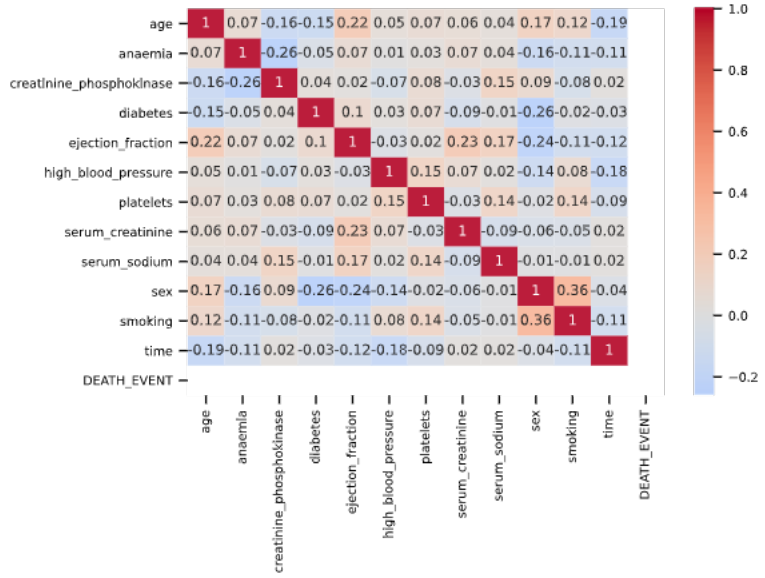
Question 1

2. for each dataset, construct the visual representations of correponding correlation matrices M0 (from df 0) and M1 (from df 1) and save the plots into two separate files

DEATH EVENT = 0



DEATH EVENT = 1



3. examine your correlation matrix plots visually and answer the following:

(a) which features have the highest correlation for surviving patients?

df_0: Sex and smoking, because |correlation|=0.49

(b) which features have the lowest correlation for surviving patients?
df_0: Sex and smoking, because |correlation|=0.01

(c) which features have the highest correlation for deceased patients?
df_1: Sex and smoking, because |correlation|=0.36

(d) which features have the lowest correlation for deceased patients?
df_1: Sex and serum sodium, because |correlation|=0.01

(e) are results the same for both cases?
No, they are not the same

Question 2

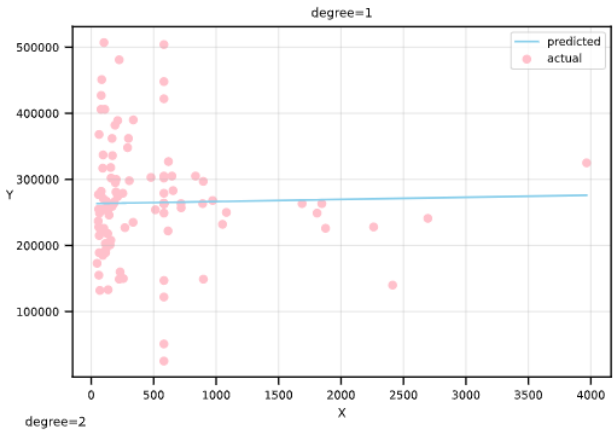
1. Group 1: X: creatinine phosphokinase (CPK), Y : platelets

(a) fit the model on Xtrain
(b) print the weights (a, b, . . .)
(c) compute predicted values using Xtest
(d) plot (if possible) predicted and actual values in Xtrain
(e) compute (and print) the corresponding loss function

Df_0

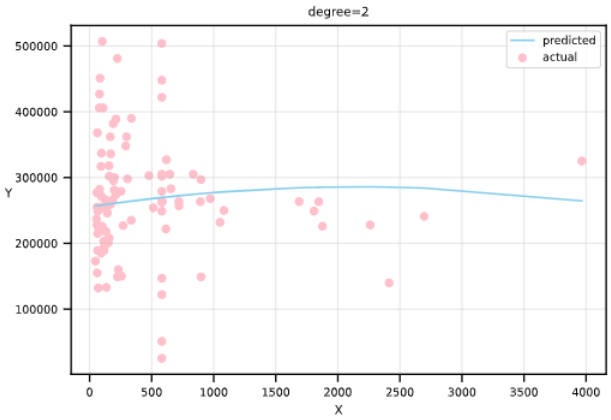
1. Y = a X + b:

weights
[3.18410860e+00
2.63325396e+05]
sse
760919264477.88



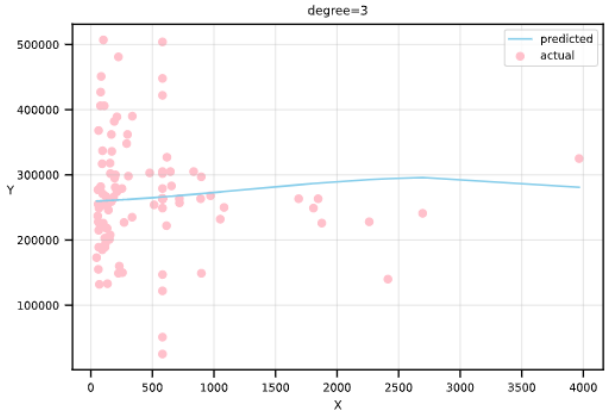
2. Y = a X**2 + b X + c

weights
[-6.51801582e-03
2.81466738e+01
2.55457684e+05]
sse
779752231349.69



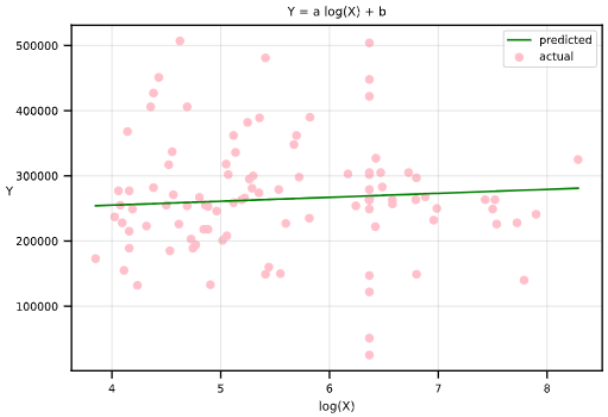
3. Y = a X**3 + b X**2 + c X + d

weights
[-0.0000000e+00
1.0000000e-02
7.5700000e+00
2.5938451e+05]
sse
780795273461.41



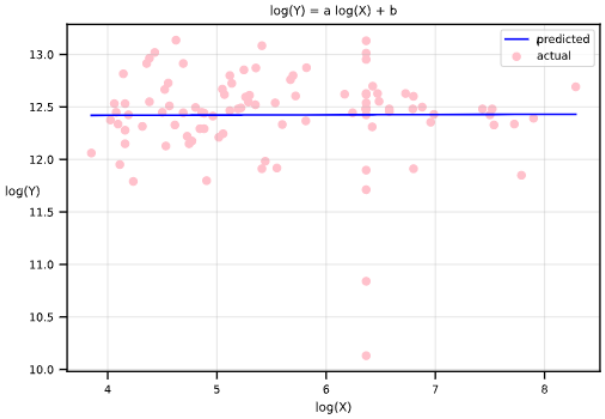
4. Y = a log(X) + b

weights
[6080.54
230655.4]
sse
768122456388.29



5. log(Y) = a log(X) + b

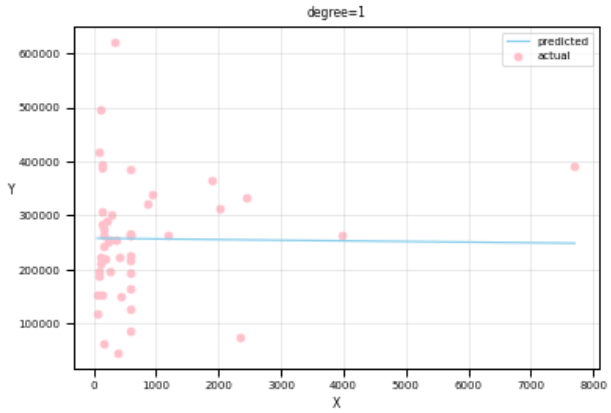
weights
[2.3000e-03
1.2411e+01]
sse
16.8902



df_1

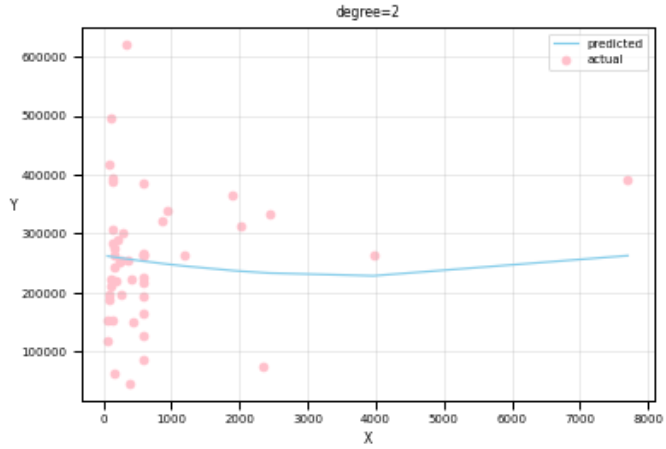
. Y = a X + b

weights
[-1.1900000e+00
2.5795916e+05]
sse
574754635199.94



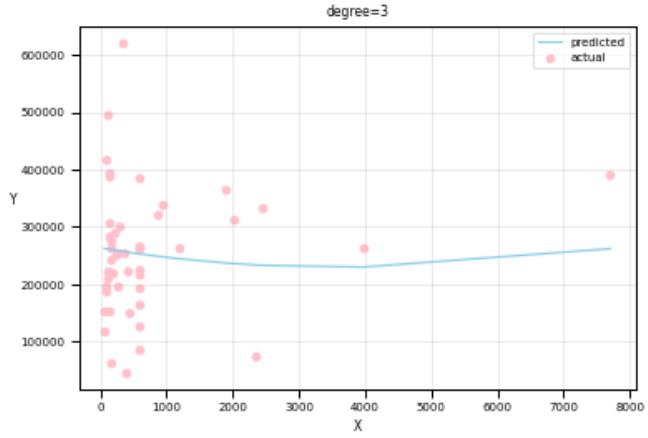
2. Y = a X**2 + b X + c

weights
[0.0000000e+00
-1.7910000e+01
2.6318718e+05]
sse
575868719642.21



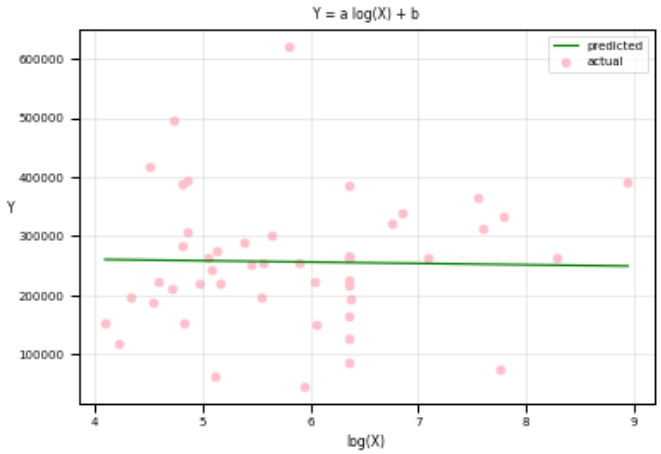
3. Y = a X**3 + b X**2 + c X + d

weights
[-0.0000000e+00
0.0000000e+00
-1.9490000e+01
2.6356626e+05]
sse
576058334804.35



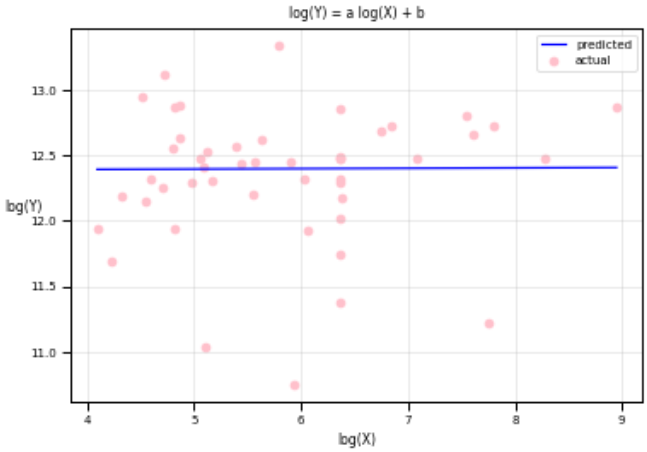
4. Y = a log(X) + b

weights
[-2294.49
270054.83]
sse
574836482570.88



5. log(Y) = a log(X) + b

weights
[2.90000e-03
1.23818e+01]
sse
12.3113



Question 3

Summarize your results from question 2 in a table like shown below:

	death event=0	death event=1
Y = a X + b	7.6092E+11	5.7475E+11
Y = a X**2 + b X + c	7.7975E+11	5.7587E+11
Y = a X**3 + b X**2 + c X + d	7.808E+11	5.7606E+11
Y = a log(X) + b	7.6812E+11	5.7484E+11
log(Y) = a log(X) + b	16.8901865	12.311272

1. which model was the best (smallest SSE) for surviving patients? for deceased patients?
for surviving patients: log(Y) = a log(X) + b
for deceased patients: log(Y) = a log(X) + b
2. which model was the worst (largest SSE) for surviving patients? for deceased patients?
for surviving patients: Y = a X**3 + b X**2 + c X + d
for deceased patients: Y = a X**3 + b X**2 + c X + d