

Marisa Paone  
MET CS677  
7/31/2023  
Homework #4

Question 1.

**Please see main.py and associated png files.**

1. *Examine your correlation matrix plots visually and answer the following:*
  - a. *Which features have the highest correlation for surviving patients?*  
**For surviving patients, the highest correlation is between serum sodium and serum creatinine: 0.22 (closest to 1 or -1).**
  - b. *Which features have the lowest correlation for surviving patients?*  
**For surviving patients, the lowest correlation is serum sodium and platelets: 0.0018 (closest to 0).**
  - c. *Which features have the highest correlation for deceased patients?*  
**For deceased patients, the highest correlation is between serum sodium and creatinine phosphokinase: 0.15.**
  - d. *Which features have the lowest correlation for deceased patients?*  
**The lowest correlation for deceased patients is serum creatinine and platelets: -0.029.**
  - e. *Are results the same for both cases?*  
**No, the results are different for both cases.**

Question 2.

**Please see main.py**

Question 3.

**Please see main.py and png files of graphs.**

1. *Summarize your results from question 2 in a table like shown below.*

Model	SSE (death_event = 0)	SSE (death_event = 1)
$y = ax+b$	1589.9289657642921	1286.2299569326603
$y = ax^2+bx+c$	1539.985467347811	1284.8596845808988
$y = ax^3+bx^2+cx+d$	1565.9644825344972	1285.3537470417343
$y = a\log x + b$	1580.573794001695	1281.1361854295956
$\log y = a\log x + b$	1593.4817471622594	1274.9715404356466

2. *Which model was the best (smallest SSE) for surviving patients? for deceased patients?*  
**The model that was the best for surviving patients was the quadratic model with an SSE of 1539.99. For deceased patients the model that was the best was the logarithmic model of  $\log y = a\log x + b$  at 1274.97.**

3. *Which model was the worst (largest SSE) for surviving patients? for deceased patients?*  
**The model that was the worst for surviving patients was the logarithmic model  $\log y = a\log x + b$  at 1580.57. For deceased patients the model that was the worst was the simple linear regression model of  $y = ax+b$  at 1286.23.**