Christopher Kruger

Midterm – Part 1

Section 1: General questions  
1. Describe how you can apply Data Science/Machine Learning to the EX05 to better analyze COVID  
comorbidities (no coding required, 1 paragraph with a description)  
2. What was the reason to create the function "print\_details" in EX04?  
3. Referring to EX05, why it is important to write a story out of the data?

1. Although difficult to pinpoint how Data Science/Machine Learning can be directly applied to Example 5, I believe that one example of Machine Learning that can be added is predictive modeling capabilities. Assuming a larger data set for added accuracy, the Machine Learning model would be able to identify trends and common factors among patients who passed. As an example, if the predictive model recognized that the highest mortality rate is among those who are >85 years old and those with diabetes, then the model can give a probability that a living patient who meets these criteria has a certain x% of passing away. This % chance of passing away would be accurately reflected by the model to be higher than say a healthy 20 year old.
2. The reason for creating the function “print\_details” was for a few reasons. One reason, and the immediate reason, is because we just learned in class how to create functions, use parameters, and how to manipulate the input into a desired output. A second reason is to make the program more modular, and to not be hard coded for use with only CSV file with specified information already known. A third reason is for reusability because it is increases processing time, file size, and makes it harder to read the code for a user if the functionality of the “print\_details” function was just copy-pasted multiple times.
3. One reason is for practicing how to interpret data and for explaining data for those who are not aware of how the algorithm works nor how to read the initial input file. This is a common case when creating solutions for business cases/requirements where those who need the data may not need nor care for a detailed explanation on how it was made. They would only need the output information for use in their own use-case. Another reason is that it is important to recognize that data is not binary nor black and white. It is not just about saying, in this case, that “those who are older are at more risk.” The narrative allows for an explanation of why this is the case, what external health complications may lead to a higher death rate, and the actual percentages of age groups who were most affected by COVID-19. It also allows for an explanation for those who are much younger to see the data and recognize that although the chances are low, it is still a possibility to pass from COVID-19.