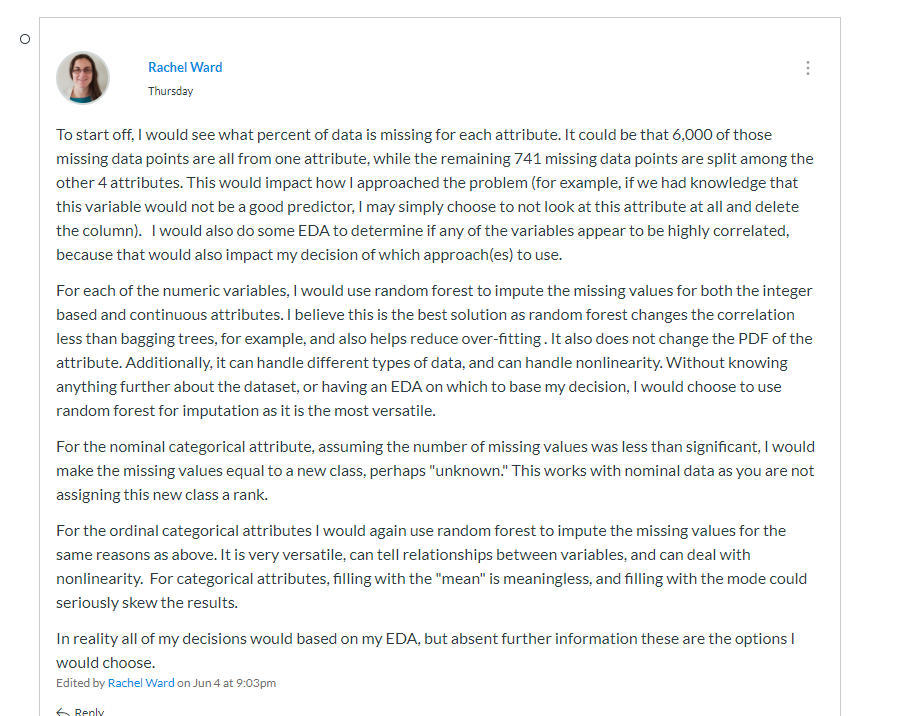
My answer:

I will use Discretization and Binning ere for the missing data values.

Since I have one continuous and one integer based values here, by using binning , I will not hide pertinent details and also not potentially introduce unnecessary "noise".

By contrast, If I delete all the value missing data, even though 6741 only in a small percentage of the totally 150,000 observations, there are still possibilities that will miss some important information.

My reply to other’s answer:



Yes, i agree with what you said Rachel. I think it would be more wise to decide which way to use for the data clean based on the EDA. And meanwhile, i think it will also necessary that we also base on the research questions that we have. For different purpose, maybe we choose choose different way to handle with the missing value observations.